The Dissertation Committee for Hsiao-Wei Liu Certifies that this is the approved version of the following dissertation:

Single-Molecule Studies on the Role of HIV-1 Nucleocapsid Protein/Nucleic Acid Interaction in the Viral Replication Cycle

Committee:

Paul F. Barbara, Supervisor

Peter J. Rossky

David A. Vanden Bout

Wolfgang Frey

Rick Russell

Single-Molecule Studies on the Role of HIV-1 Nucleocapsid Protein/Nucleic Acid Interaction in the Viral Replication Cycle

by

Hsiao-Wei Liu, B.S. Ch.

Dissertation

Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

The University of Texas at Austin
August 2007

Dedication

我將這本畢業論文獻給我的父親及母親

Acknowledgements

I couldn't ask for a better environment than Dr. Barbara's lab for me to do my Ph.D. The Barbara group consists of knowledgeable postdoctoral fellows coming from a variety of scientific training. I appreciated that each of them was willing to take time to help me with the experimentally and curricularly scientific questions. I was also lucky to be part of history of NC/HIV-1 project. Working with Dr. Cosa (currently as a professor in McGill University) and Mr. Zeng in the infant period of the project, we constantly encountered obstacles and failed measurements. Later, I realized how valuable these failed trials are because so often when the experiment is working fine, I took everything for granted and forgot to observe, think and ask questions. In addition, to be able to work for Dr. Barbara was truly amazing. Under his guidance, not only my research went very well but also the training he gave me made me a better scientist. So often, I attended his talks and was 'infected' by his enthusiasm towards the research and realized that the research I completed was inspiring. I also learned from him that as a scientist I should never stop learning the science. Five years has passed now. I feel ready to move on and start my postdoctoral training. From the bottom of my heart, I am thankful for those who have helped me through many obstacles during my Ph.D. studies and of course my parents for their endless support.

Single-Molecule Studies on the Role of HIV-1 Nucleocapsid Protein/Nucleic Acid Interactions in Viral Replication Cycle

Publication No._____________

Hsiao-Wei Liu, Ph.D.
The University of Texas at Austin, 2007

Supervisor: Paul F. Barbara

The discovery of the crucial intermediates and pathway in the process of the reverse transcription was reported using single-molecule spectroscopy and related techniques including single-molecule fluorescence resonance energy transfer, fluorescence correlation spectroscopy and confocal imaging. Reverse transcription of the HIV-1 RNA genome involves several complex nucleic acid rearrangement steps that are catalyzed by the HIV-1 nucleocapsid protein (NC), including for example, the annealing of the transactivation response (TAR) region of the viral RNA to the complementary region (TAR DNA) in minus-strand strong-stop DNA. In this dissertation, the research focused on elucidating the mechanism of NC-facilitated TAR DNA/RNA annealing. The single molecule spectroscopic measurements reported that the crucial intermediates as well as the mechanistic insight into the annealing of TAR RNA with TAR DNA mediated by viral NC proteins. The data reveal that NC partially melted the secondary structure of TAR DNA (termed the "YTAR") as well as TAR RNA. In the subsequent studies, various short DNA oligonucleotdies were applied to anneal with the TAR to mimic the initial annealing steps. The data support that the YTAR serves as a nucleation center for the annealing to occur through the multiple sites along the TAR structure. Two major nucleation pathways were observed, which are the annealing through the 3'/5' termini, namely "zipper" pathway and the annealing through the hairpin loop region, namely "kissing" pathway. The annealing mechanism was further explored by performing the annealing of wild-type TAR DNA with wild-type TAR RNA in the presence of NC *in vitro*. The annealing kinetic data suggest that the nucleation of TAR DNA/RNA annealing occurs in an encounter complex form in which one or two DNA/RNA strands in the "Y" form associated with multiple NC molecules. This encounter complex leads to the multiple nucleation complexes, i.e. zipper or kissing intermediates. The data further indicate that although the two complementary strands nucleate at multiple sites, i.e. any single-strand region of TAR, the annealing of two TAR complements occurs through a common mechanism.

Table of Contents

List of Tables

List of Figures

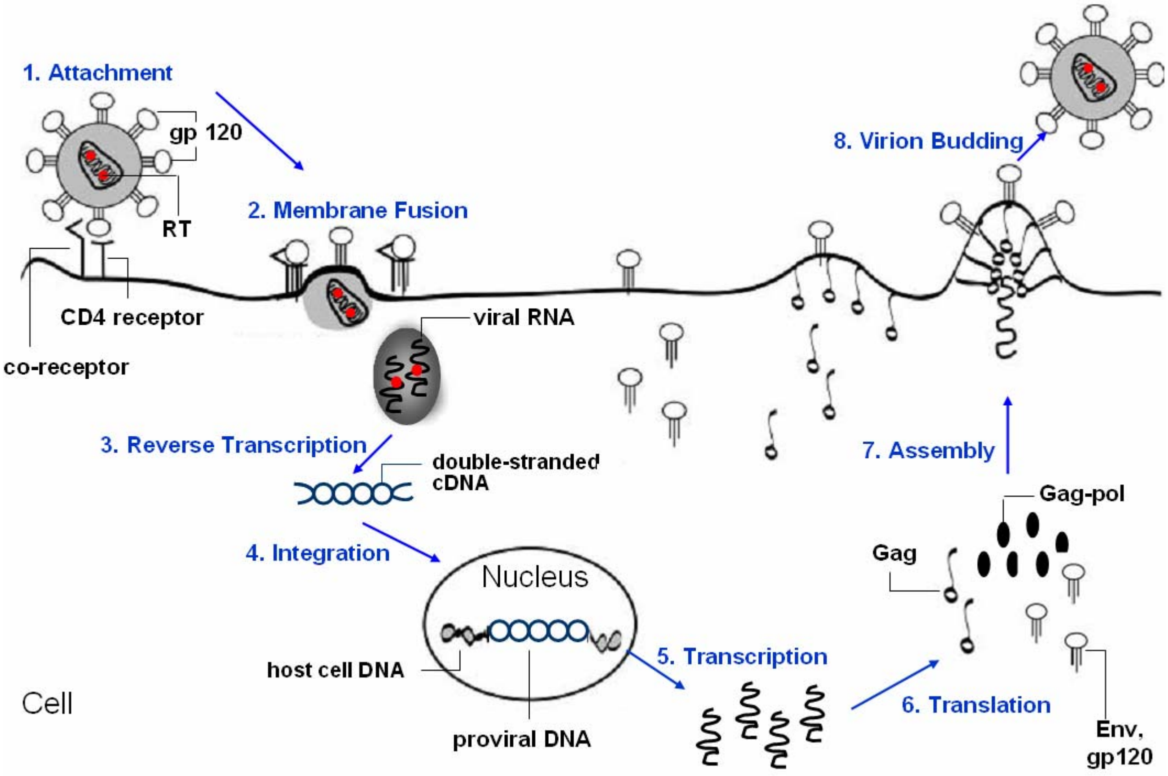
List of Schemes

Chapter 1. Introduction

1.1 THE HIV-1 VIRUS LIFE CYCLE

The HIV-1 virus is a retrovirus consisting of two identical single-stranded RNA genomes encapsulated by various viral structure proteins and exhibits a cone-shaped capsid core particle. On the shell of the viral particles (or virions), many copies of envelope proteins are embedded in the lipid bilayer of the viral envelope. Scheme 1.1 depicts a simplified picture of the HIV-1 virus life cycle[1]. The infection of HIV-1 viruses is initiated by the envelope gp120 glycoprotein on the surface of the virion recognizing CD4 receptors on a T lymphocyte (1, Scheme 1.1). Following the recognition of CD4 receptors, the virions fuse into the cells (2, Scheme 1.1). The core viral proteins and diploid RNA are disassembled and the RNA genomes are released into the cytoplasm. Subsequently, the viral single-stranded RNA genomes are converted into a double-stranded DNA (referred to as cDNA) by a viral protein, reverse transcriptase (RT) (3, Scheme 1.1, for the detailed description of reverse transcription, please see Section 1.2). Thereafter one or more copies of cDNA are transported into the cellular nucleus and integrated into the host chromosome by another viral protein, integrase (IN) (4, Scheme 1.1. The integrated viral DNA (known as a proviral DNA) then serves as a template transcribed into viral RNA which then serves as mRNA translated to various viral accessory proteins including Gag proteins in many cycles (5 and 6, Scheme 1.1). Shortly after the translation, thousands of newly synthesized Gag proteins accumulate on the cytoplasmic face of the plasma membrane driving the viral particle assembly through Gag-Gag interactions (7, Scheme 1.1). By forcing the membrane into curvature, the virion assembly protrudes from and pinches off the cell ultimately leading to the release of the virions (referred to as virion budding, 8, Scheme

1.1). It has been shown that HIV-1 viral replication occurs at a very rapid rate, producing over 1-10 billion virions per day in an infected body[2], and meanwhile destroying a similar number of T lymphocytes in the process[3].

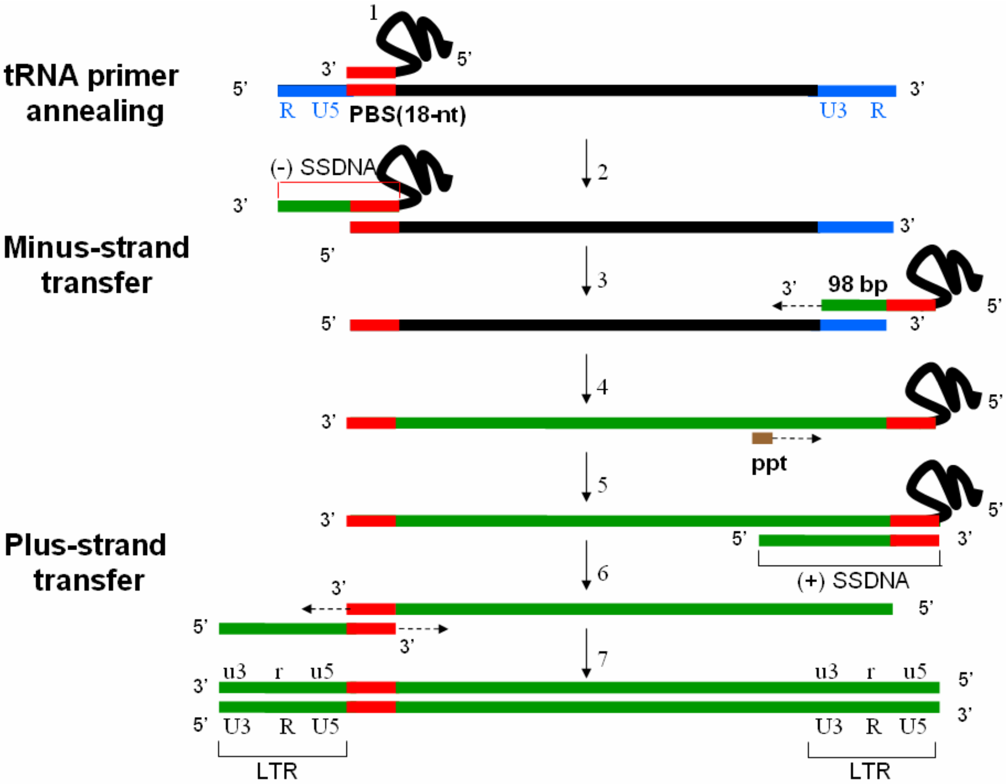


Scheme 1.1: The life cycle of HIV-1 virus.
This scheme is adapted from G. Cristofari and J.-L. Darlix, 2002[1].

1.2 REVERSE TRANSCRIPTION

Reverse transcription is a critical step for retroviral replication[4]. The conversion of a single-stranded RNA genome to a double-stranded DNA is achieved by employing two obligatory strand transfer steps, minus-strand transfer and plus-strand transfer. During the assembly of viruses, both host $tRNA^{Lys,3}$ and viral RNA are selectively packaged into the virions.

The HIV-1 reverse transcriptase initiates proviral DNA synthesis from a tRNA^{Lys,3} annealed to an 18-base sequence of the primer binding site (PBS) near the 5' end of the RNA (step 1, Scheme 1.2). Upon the hybridization of tRNA^{Lys,3} to PBS, the polymerase domain of viral RT copies the RNA into DNA sequences, and simultaneously the RNase H domain of RT degrades the genomic RNA which has been copied (step 2, Scheme 1.2). Synthesis of the DNA copy proceeding to the 5' end of the genomic RNA generates the minus-strand strong stop DNA ((-) SSDNA).



Scheme 1.2: Mechanism of reverse transcription.

Step 1: Annealing of tRNA^{Lys,3} to the PBS. Step 2: Synthesis of (-) SSDNA and digest of RNA template by RT. Step 3: Minus-strand transfer. Step 4: Elongation of minus-strand DNA to the 3' end. Step 5: Synthesis of (+) SSDNA initiated from PPT as a primer. Step 6: Plus-strand transfer. Step 7: Elongation of both DNA strands. Abbreviations: U5/3, 5'/3' untranslated region; PBS, primer binding site; SSDNA, strong stop DNA; PPT, polypurine tract; LTR, long terminal repeat region.

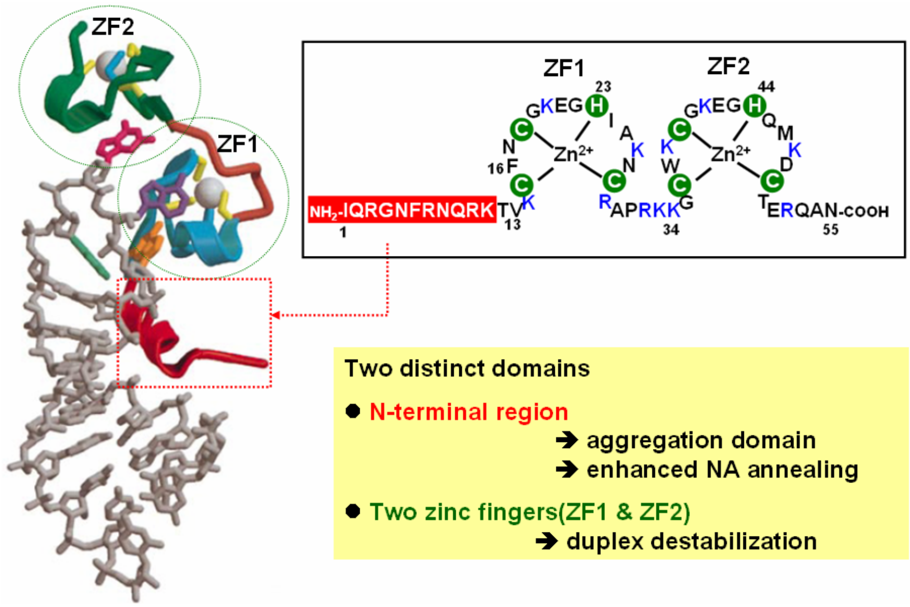
In order to complete the reverse transcription, (-) SSDNA must be transferred to the 3’ end of the RNA genome. Since the genomic RNA has two identical R regions in the terminus, base-pairing (-) SSDNA (containing partially sequences of the R region) with the RNA genome can be switched from the 5’ end to the 3’ end (step 3, Scheme 1.2). This so-called minus-strand transfer step is obligatory and is either intra-molecular (i.e. from one end to the other end of a genomic RNA) or inter-molecular (i.e. from one to the other copy of genomic RNA)[5]. After the minus-strand transfer, RT continues to elongate the same DNA transcript until it reaches the polypurine tract (PPT) site located near the 3’ end of the viral genome (step 4, Scheme 1.2). The RNase H domain of RT makes a specific nick at the PPT of 5’ end of RNA genome and creates an RNA primer to initiate plus-strand DNA synthesis (step 5, Scheme 1.2)[6,7].

Once the synthesis of plus-strand DNA reaches the 5’ end of minus-strand DNA, plus-strand strong-stop DNA is then transferred to the other end of the almost complete minus-strand DNA, presumably by the complementarity between the PBS (step 6, Scheme 1.2). This obligatory plus-strand transfer is primarily intra-molecular and enables the elongation of plus-strand DNA to proceed to the 3’ terminus. Two obligatory transfers result in extension of both DNA transcripts and create the long terminal repeat regions (LTR) containing extra untranslated regions that do not exit in the viral RNA at both termini (step 7, Scheme 1.2).

1.3 THE HIV-1 NUCLEOCAPSID PROTEINS

1.3.1 The Nucleic-Acid-Chaperone Activity of Nucleocapsid Proteins

Shortly subsequent to virion budding, the virion undergoes a major structural change termed ‘maturation’ in which the Gag polyprotein is cleaved by the virus-encoded protease (PR) into a series of products, including nucleocapsid (NC) proteins. The HIV-1 NC is a zinc-finger protein with 55 amino acids. NC is composed of two zinc-binding domains of the CCHC type flanked by a flexible polypeptide chain and connected by a short basic peptide linker (see Scheme 3.1, inset)[8-11]. About 2000 copies of NC are tightly associated with the genomic RNA in the core of mature virions[12,13]and are believed to protect the genomic RNA against degradation by RNase[14].



Scheme 1.3: Three-dimensional NMR structure of a NC/RNA hairpin complex.

The N-terminal 12 residues of NC (shown in red) bind to the major groove of the RNA stem (shown in gray) through the electrostatic interactions. The two zinc fingers (ZF1 and ZF2, shown in blue and green) engage in specific interactions three purine bases (shown in green, pink, purple and orange) in the tetraloop. The zinc ions are shown in white. (Inset) The primary structure of NC protein shown is from the NL4-3 isolate of HIV-1. The cysteine and histidine residues coordinating with zinc highlighted in green show both zinc fingers belonged to the “CCHC” type. The basic residues are shown in blue. This figure is adapted from B. G. Turner and M. F. Summers, 1999[13].

The function of NC relies on two interactions, electrostatic and π- π stacking interactions, with nucleic acids (NAs). The N-terminal basic 3_{10} helix of NC interacts nonspecifically with the major groove of the RNA stem by electrostatic interactions[15,16] and is responsible for the aggregation of NC/nucleic acid complexes[17,18]. The zinc finger domain interacts with specific base sequences in particular with a preference for TG- or UG- rich regions[19-21]. The binding of zinc fingers to NA is responsible for destabilizing NA helix[22]. NMR studies further showed that Phenylalanine in the second zinc finger (ZF1) and Tryptophan in the second zinc finger (ZF2) engaged in aromatic π-π stacking interaction with the G residues in the single-stranded GGAG tetraloop as shown in Scheme 1.3[13]. Although in these cases NC was observed to have a preferential binding to single-stranded sequences[19-21,23-26] with a binding site of 5-8 nucleotides (nt)[16,19,27-32], binding to single- and double-stranded sequences with comparable affinities was also observed[25]. The binding of NC has been shown to be salt-dependent and is driven apparently by the release of Na^+ or other counter ions[16,32,33].

As a nucleic acid chaperone protein, NC catalyzes NA conformational rearrangements that lead to the most thermodynamically stable structure[1,34-37]. The chaperone activity of NC is believed to be derived from two main consequences of the nucleic-acid/protein (NC) interactions in this system. First, NC lowers the energy of highly structured nucleic acids by the complexation with nucleic acids and further partially melting NAs. Second, NC lowers the energy cost of bringing two complementary strands of NAs together by screening the negative charges of two NAs[17,18,31,38,39] and perhaps through the binding of the N-terminal basic 3_{10} helix to the double-stranded stem of the hairpins.

1.3.2 The Role of Nucleocapsid Proteins in the HIV-1 Viral Life Cycle

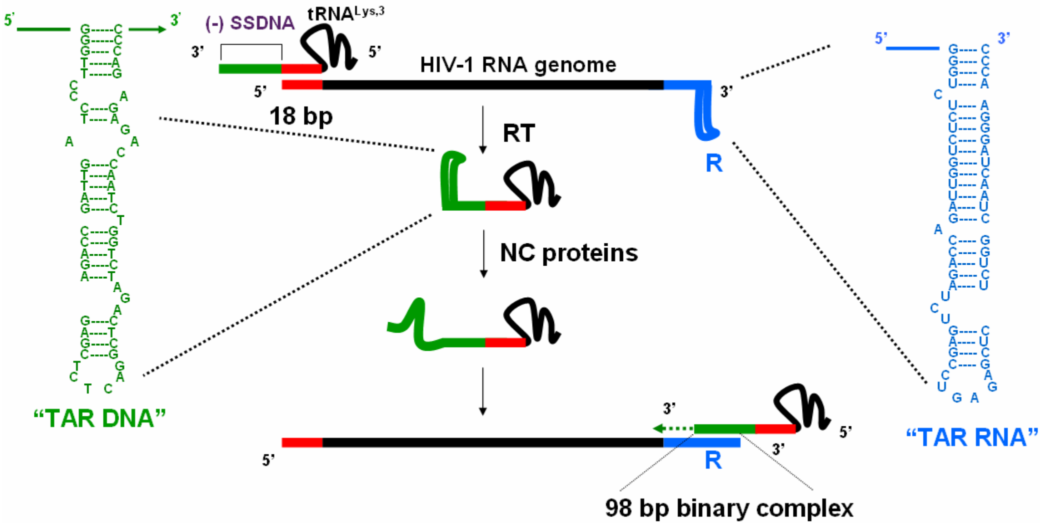
The NC (or the NC domain of Gag precursor) is crucial for many processes of the viral life cycle in which interactions between nucleic acid and NC are the key. For example, as a domain of Gag proteins, NC is involved in the recognition and packaging of RNA genomes [40-42], in virus assembly[43], in selectively packaging tRNA^{Lys,3} primer [44] and in many annealing steps during the reverse transcription[45-52]. The NC domain of Gag proteins is also important for viral particle assembly[53,54], presumably through NC-NC interactions[13,55].

The chaperone activity also allows NC to be a cofactor of RT and promote several annealing reactions during the HIV-1 reverse transcription process. For example, NC makes it possible for RT to copy the stem-loop sequences in RNA and DNA templates by melting the structured sequences[56]. In additions, NC has been shown to facilitate the annealing of cellular tRNA onto the PBS of the viral RNA[46,57], the annealing of complementary PBS sequences in plus-strand transfer[58-60] and the annealing of the transactivation response (TAR) region of the viral RNA to the complementary sequences (TAR DNA) during minus-strand transfer[59,61].

1.3.3 The Role of Nucleocapsid Proteins in the Minus-Strand Transfer

During the minus-strand transfer, (-) SSDNA is transferred from the 5' to the 3' end of the RNA genome by base-pairing of the complementary repeat regions in (-) SSDNA and viral RNA. The result of minus-strand transfer is for DNA/RNA hybrid to lose 18 base-pairs and to gain 98 base-pair. Although thermodynamically favored, this reaction does not occur significantly in the absence of NC (3% without NC and 65% with NC)[62-66]. This reaction involves annealing of the highly stable TAR DNA in (-) SSDNA with its complementary TAR RNA present in the viral RNA (see scheme 1.4). Therefore, kinetically it is difficult to form the final product of RNA/DNA hybrid.

Several groups have shown that NC promotes minus-strand transfer due to its ability to unfold the TAR DNA/RNA hairpin with stem-loop structures in the R regions[67-70]. The unfolding of these hairpins induced by NC was thought to be rate-limiting[61]. In addition, absorbance measurements showed that NC had a greater effect on melting of TAR DNA than on melting of TAR RNA[71] presumably due to the greater stability of the RNA hairpin. These stable *intramolecular* structures also prevent the *intermolecular* annealing reaction that results in the formation of a final product of minus-strand transfer with a 98-nt base paired binary complex.



Scheme 1.4: Minus-transfer step during the reverse transcription catalyzed by NC.

NC is shown to destabilize the highly structured TAR RNA (shown in blue) and TAR DNA (shown in green) and to promote the annealing of two hairpins leading to the formation of a 98-nt base pair duplex.

NC has been shown to stimulate minus-strand transfer by increasing the rate of intermolecular annealing by as much as 3,000-fold and by blocking a competing intramolecular self-priming reaction[61,66,72,73]. The self-priming occurs due to the presence of the TAR DNA hairpin at the 3′ end of the (-)SSDNA initiating unwanted elongation by-product[4,61,65,66].

1.4 SINGLE-MOLECULE APPROACHES TO UNRAVELING THE TAR DNA/RNA ANNEALING

1.4.1 Single-Molecule Spectroscopic Approaches to Biological System

Two most popular single-molecule spectroscopy (SMS) techniques applied in biological studies are Single-molecule Florescence Resonance Energy Transfer (SM-FRET) and Fluorescence Correlation Spectroscopy (FCS). SM-FRET measurement for biological applications was first performed in 1996[74] while the concept of FCS was already introduced in the early 1970s[75]. SM-FRET, a fairly new technique as compared to FCS, has been extensively applied for the direct observation of the structural dynamics of proteins, protein-protein interaction[76,77] and DNA–protein interactions *in vitro*[67,78-80] in the past decade. Likewise, the recent FCS technological advances have rendered FCS one of the most powerful tools as SM-FRET for biological research. Fluorescence fluctuations due to concentration fluctuations are readily detected yielding the information of molecular diffusion, interaction between molecules and intramolecular protein dynamics[81].

These single-molecule spectroscopic techniques allow one to observe one molecule at a time and resolve subpopulations in heterogeneous samples that would otherwise be hidden in ensemble averaging process. Single-molecule approaches also reveal transient intermediates and temporal heterogeneous behavior by recording asynchronous time trajectories. The ability of single-molecule spectroscopy to reveal intricate biomolecular dynamics thus renders this technique particularly powerful in elucidating the mechanisms of molecular machineries.

1.4.2 Single-Molecule Fluorescence Resonance Energy Transfer Spectroscopy

Florescence Resonance Energy Transfer (FRET) is based on the theory that electronic excitation energy transfer can efficiently occur via dipole-dipole interactions from a donor dye to an acceptor dye in a distance-dependent manner. The FRET efficiency (E_{FRET}) is a direct measure of the fraction of photon energy absorbed by the donor transferred to an acceptor. E_{FRET} can be calculated as the ratio of the energy transfer rate k_T to the total decay rate of the donor as Equation 1.1:

$$E_{FRET} = \frac{k_T}{{\tau_D}^{-1} + k_T} \tag{1.1}$$

where τ_D is the lifetime of the donor in the absence of the acceptor.

Experimentally E_{FRET} is measured as the relative fluorescence of the donor in the presence (I_{DA}) and in the absence (I_D) of the acceptor as Equation 1.2 equal to Equation 1.3 as an approximation where I_A is the fluorescence of the acceptor caused by the energy transfer from the donor[82].

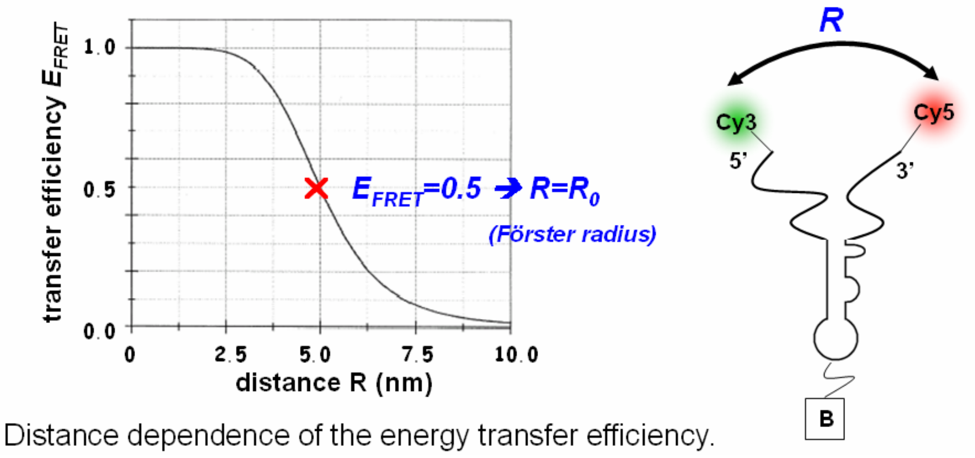
$$E_{FRET} = 1 - \frac{I_{DA}}{I_D} \tag{1.2}$$

or $$E_{FRET} = \frac{I_A}{I_A + I_D} \tag{1.3}$$

E_{FRET} also depends on the inverse of the sixth power of the distance R between the two fluorophores as:

$$E_{FRET} = \frac{1}{1 + (\frac{R}{R_0})^6} \tag{1.4}$$

where R_0 is known as the Förster distance and is a characteristic parameter of each donor-acceptor pair defining the distance at E_{FRET}=0.5 (see Scheme 1.5, *left*). Depending on the labeling configuration, E_{FRET} can be associated with either intra- or inter- changes in the distance of a donor-acceptor pair further indicating the structural changes of the dye-labeled species. In the case of single-molecule FRET, dye-labeled molecules are immobilized on the surface of the coverslips (see Scheme 1.5, *right*). By taking time traces of the donor and acceptor fluorescence signals from individual molecules, the FRET of a single molecule can be calculated from Equation 1.3 (see Section 2.3 for the experimental details).



Scheme 1.5: An example of SM-FRET measurements.

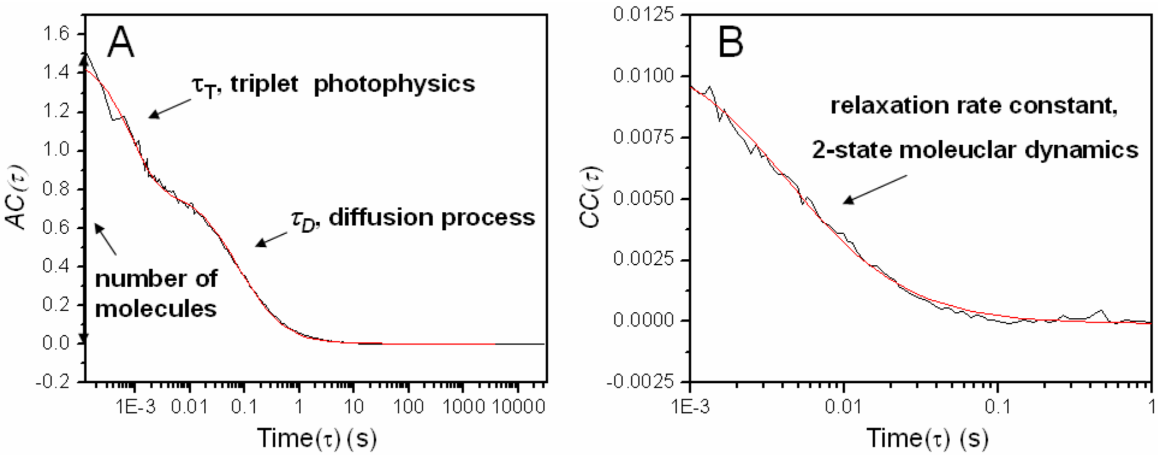
Left: Distance dependence of E_{FRET} is shown using Cy3-Cy5 as a FRET pair. The R_0 for Cy3-Cy5 is 5 nm. *Right*: SM-FRET measurement can be performed by using an intra-molecular FRET pair attached to a biotynlated single molecule DNA hairpin.

1.4.3 Fluorescence Correlation Spectroscopy

Fluorescence correlation spectroscopy (FCS) is a spectroscopic technique for analyzing the correlation of fluctuations of fluorescence intensity. The fluorescence intensity fluctuates due to the fluctuations of number of the emitting molecules (usually around nM) inside a tiny volume irradiated by a focused laser beam (focal volume, ~1 fL) with temporal resolution typically from 1μs to >10s. The correlation analysis can be performed by auto-correlating the fluctuation of a fluorescence signal *I(t)* (autocorrelation, *AC(τ)*) defined by

$$AC(\tau) = \frac{<\delta\ I(t)\delta\ I(t+\tau)>}{<I(t)>^2} = \frac{<I(t)I(t+\tau)>}{<I(t)>^2} - 1 \qquad (1.5)$$

where *<I(t)>* is the mean of *I(t)* and *δI(t)* is the difference of *I(t)* from *<I(t)>*. Here the *AC(τ)* is represented in its normalized form in which the autocorrelation is divided by the mean *<I(t)>*^{2}.



Scheme 1.6: Two types of FCS measurements.

(A) Autocorrelation curve of molecules diffusing into the focal volume provides triplet dynamics and diffusion coefficient of fluorophores. (B) Auto-correlation curve of a molecule labeled with a FRET pair provides the relaxation rate constant of a two-state-dynamics.

In a model of fluorophores freely diffuse into a 3D Gaussian focal volume with contributions from triplet photophysics, the correlation function can be expressed as follows[83]:

$$G(\tau) = 1 + \frac{1}{N}(1+\frac{\tau}{\tau_D})^{-1}(1+\frac{\tau}{\omega^2\tau_D})^{-1/2}[1+\frac{p}{1-p}\exp(-\frac{\tau}{\tau_T})] \qquad (1.6)$$

In the equation 1.6, the term N is the number of molecules passing through the excitation volume and τ_D is the diffusion time of the molecule within this excitation volume. ω =

ω_z/ω_{xy} is the height-diameter ratio of the 3D Gaussian focal volume of the laser with the Gaussian width of ω_z and ω_{xy} in axial and planar directions, respectively. p is the fraction of the molecules in the triplet state, and τ_T is the relaxation time of the triplet state. The value of $G(\tau)$ at $\tau = 0$, which is inversely proportional to N determines the average number of fluorophores. This FCS analysis yields the average number of the emitting molecules and average diffusion time when the molecules are passing through the focal volume (see Scheme 1.6A). This information is used to further determine both the concentration and mass of the molecules.

In a modal of a donor-acceptor labeled molecule fluctuating between two states as Equation 1.7, SM-FRET measurements record the signal fluctuation of donor or acceptor dyes at two states with a forward and backward rate constant.

$$state\ 1 \underset{k_{backward}}{\overset{k_{forward}}{\rightleftharpoons}} state\ 2 \tag{1.7}$$

The $AC(\tau)$ of fluorescence signals between two states, I_A^1 (or I_D^1) and I_A^2 (I_D^2) in this case can be expressed as[84].

$$AC(\tau) = \frac{(I_A^{\ 2} - I_A^{\ 1})^2 k_{forward} k_{backward}}{k_{forward} I_A^{\ 1} + k_{backward} I_A^{\ 2}} e^{-\lambda t} \tag{1.8}$$

where $\lambda = k_{forward} + k_{backward}$ is a relaxation rate constant. The relaxation rate constant extracted from the calculated auto-correlation provides the time scale for dynamics of a two-state system (see Scheme 1.6B, for the experimental details of FCS please see Section 2.4).

1.4.4 New Mechanistic Insight of TAR DNA/RNA Annealing Using Single-Molecule Approaches

Since there are only two copies of RNA genome in the viral core, each viral RNA is associated with thousands of NC and each segment of the RNA genome interacts with locally high concentrations of NC. Single-molecule methods allow the NC/nucleic acid interactions to be studied at *in vivo* ‘native’ concentrations. In addition, the NC proteins are prone to aggregate at ensemble concentration (> μM) due to the aggregation domains of NC[17]. Therefore, single-molecule methods operating at low concentrations (nM scale) allows the mechanism of TAR DNA/RNA annealing monitored without the aggregation complications.

SMS approaches including SM-FRET, FCS and confocal imaging, have brought more mechanistic insights into NC-mediated TAR RNA/DNA annealing. SM-FRET studies combined with FCS measurements on single immobilized wild type (WT) and mutant TAR DNA, derived from the NL4-3 isolate, complexed with NC (i.e. TAR DNA/NC) reported the conformational distribution and dynamics of folding/unfolding of the hairpins in the presence and absence of NC protein. This was the first measurement to identify the structure of unfolding TAR induced by NC. The data demonstrate that NC shifts the equilibrium secondary structure of WT TAR DNA hairpins from a fully “closed” conformation to one specific partially-open “Y” conformation[67]. This Y conformation is arguably a key TAR DNA intermediate in the NC induced annealing mechanism of TAR DNA. The information of the unfolding structure of the TAR DNA induced by NC is detailed in Chapter 3.

SM-FRET and time resolved confocal imaging measurements were also performed to investigate the annealing mechanism of TAR DNA/RNA at the single-molecule level. We examined two proposed annealing mechanisms, the so-called “zipper” and “kissing” pathways, as well as intermediates of NC-facilitated TAR RNA/DNA annealing. In the zipper mechanism, nucleation occurs through either the 3’ or 5’ terminus of TAR DNA and RNA hairpins with NC-induced “Y” structures observed by SM-FRET measurements. In the kissing mechanism, nucleation occurs through base-pairing of the single-stranded hairpin loop regions of the TAR RNA and DNA hairpins[85,86]. To examine both pathways, the single-molecule annealing experiments were perform using TAR DNA and various DNA/RNA oligonucleotides. The sequence of oligonucleotide was chosen for its specific complementarity to regions of TAR DNA characteristic for each pathway under consideration. The data showed that although formation of both zipper and kissing intermediates was observed, the zipper pathway dominates in the annealing reaction. These studies suggest that multiple annealing pathways can occur for the NC-mediated TAR RNA/DNA annealing step of minus-strand transfer. The information related to the observed zipper and kissing pathway is detailed in Chapter 4.

The zipper and kissing mechanisms were further examined by SM-FRET using full-length WT and mutant TAR DNA and RNA. The data showed that NC-facilitated TAR annealing required two terminal loops of TAR DNA/RNA. These weakly terminal bulge regions have been shown to be critical for NC-induced melting of TAR DNA. Moreover, the data suggests that the nucleation complex of the annealing reaction goes through two DNA/RNA molecules, with one or both of the molecules in the “Y” conformation coated by multiple NC molecules. The complete scheme of TAR DNA/RNA annealing promoted by NC is detailed in Chapter 5.

1.5 SIGNIFICANCE

The fact that NC functions at multiple stages during the HIV-1 life cycle makes NC an attractive target for the design of anti-HIV compounds. Mutations in the Cys and His residues of the HIV-1 NC zinc-finger domains caused major defects in the viral RNA encapsidation[40,41,54,87-89]. Mutations in the N-terminal basic 3_{10} helix also reduced RNA binding in vitro [90] and RNA encapsidation into virions[43,91]. The high mutation resistance of NC is crucial for a successful anti-HIV drug since the current three major categories of anti-HIV drugs, RT-, IN- and PR-inhibitors on the markets all exhibit drug resistance to some extents due to enzyme mutation[13].

The highly conserved CCHC type of zinc-finger motif is critical for the function of HIV-1 NC. Such a zinc-finger protein with CCHC type is relatively distinct among the cellular zinc-finger proteins, which typically contain CCHH or CCCC motifs[92]. Therefore it is possible to develop a compound which specifically targets the zinc fingers of HIV-1 NC without blocking the function of their cellular counterparts. In fact, a number of compounds were shown to disrupt NC's function by causing the ejection of zinc ions[93-95]. This inactivation is irreversible due to the formation of inter-and intramolecular disulfide crosslinks between Cys residues of the zinc-fingers following zinc ejection.

While multi-drug therapy has delayed the onset of clinical AIDS and death, there continues to be a need for anti-HIV drug for curing AIDS. ADA has been the first anti-NC drug shown to inhibit HIV-1 infection and has achieved a phase 1 clinical trial in Europe. Inactivation of virus particles by chemical attack on the zinc finger of NC is a promising approach for ultimate killed-virus vaccine development. Meanwhile, a detailed understanding of the NC-facilitated minus-strand transfer step of reverse transcription at the molecular level will contribute to developing more effective anti-viral medications and therapies.

Chapter 2. Experimental Methods

The goal of this chapter is to describe in general detail the experimental apparatuses and methods required to obtain the FRET-based fluorescence microscopic data as well as the data analysis. The apparatus and methods outlined here will be relevant for subsequent chapters that briefly highlight specific conditions for the corresponding experiments.

2.1 SAMPLE PREPARATION

2.1.1 Coverslip Surface and Flow Cell Preparation

2.1.1.1 Bovine Serum Albumin Surface

Prior to surface modification, coverslips were cleaned rigorously using the following procedure:

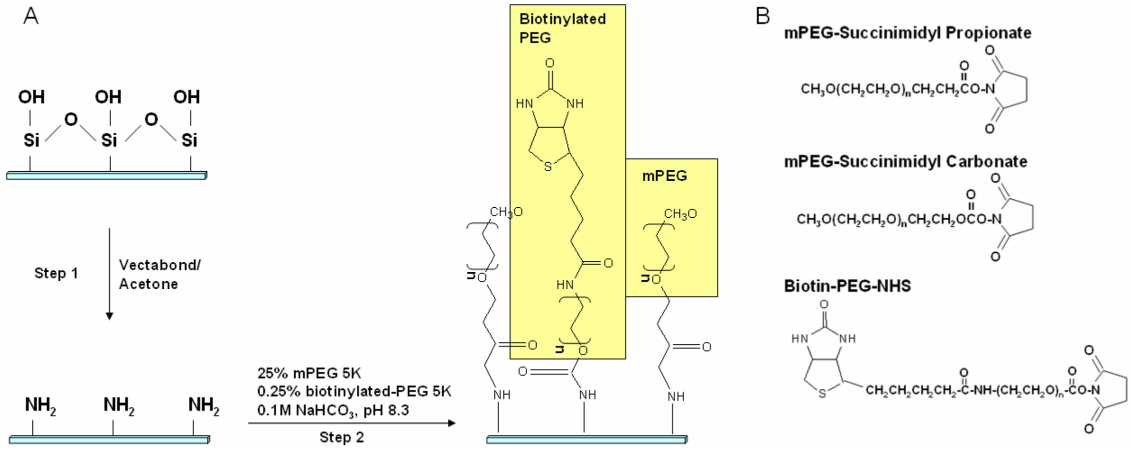
Coverslips were soaked in piranha solutions (25% H_2O_2 and 75% concentrated H_2SO_4) for 1 hr, followed by various water (molecular biology grade (MBG), HyClone, Logan, UT), and acetone (high-performance liquid chromatography (HPLC) grade, Fisher Scientific) rinsing cycles. Coverslips were dried under a N_2 stream and ~~ready~~ immediately used for flow cell assembly.

Predrilled polycarbonate films with an adhesive gasket (Grace Bio-Labs, Bend, OR) were assembled on top of cleaned coverslips yielding a flow cell with a total volume of 10 μL. Inlet and outlet ports (Nanoport, Upchurch Scientific, Oak Harbor, WA) were glued on top of the cells for the surface modification.

The bovine serum albumin (BSA) surface was prepared by incubating biotinylated BSA (Pierce Biotechnology, Rockport, IL; 2 mg/mL in MBG water) for 10 min in the flow cell and excess BSA was then rinsed out. Through non-specific adsorption the coverslip surface were biotin-functionalized for DNA immobilization.

2.1.1.2 Polyethylene Glycol Surface

Prior to polyethylene glycol (PEG) modification, coverslips were cleaned rigorously as previously described (see Section 2.1.1.1). Dry, clean coverslips were then treated with an amino-silane reagent, 1% w/v vectabond/acetone solution (Vector Laboratories, Burlingame, CA) for 5 min and then rinsed with H_2O and dried under an N_2 stream.



Scheme 2.1: The functionalized PEG surface and PEG reagents for making functionalized PEG surfaces.

The clean coverslips were masked with patterned silicone films. The unprotected area of the coverslips was incubated with a 25% w/w PEG solution (mPEG-Succinimidyl Propionate, MW 2k or 5k Da, Nektar Therapeutics, Huntsville, AL or mPEG-Succinimidyl Carbonate, MW 5k Da, Laysan Bio, Inc., Arab, AL) containing 0.25% w/w biotinylated PEG (Nektar biotin-PEG-NHS, MW 5k Da) in a 0.1 M sodium bicarbonate solution (HyClone 75.0 g/L) for 3 hr (see Scheme 2.1 B for the chemical structures of the PEG reagents). The reactive PEG surface areas were marked and the silicone films were removed. The excess PEG was rinsed with water, and the coverslips dried under an N_2 stream for flow cell assembly as previously described in Section 2.1.1.1. The functionalized PEG surface is shown below in Scheme 2.1 A.

2.1.2 Immobilization Scheme and Immobilization DNA Sample Preparation

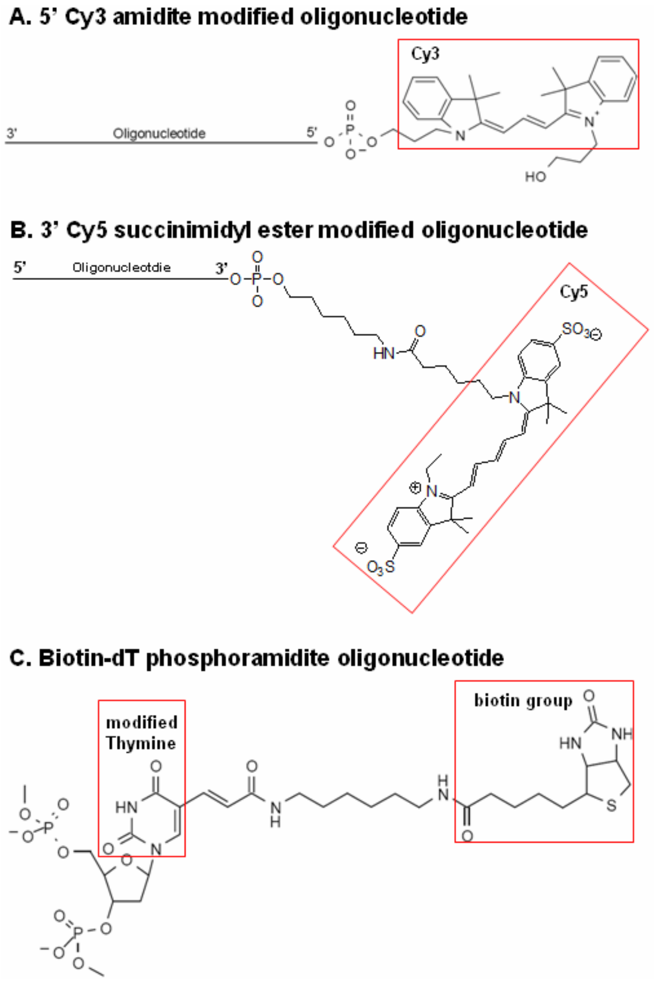
2.1.2.1 DNA Fictionalization

All oligonucleotides were purchased in a lyophilized form (TriLink Biotechnologies, San Diego, CA). The oligonucleotides were purified by the supplier by poly-acrylamide gel electrophoresis and reversed phase HPLC. The primary and secondary structures of the DNA oligonucleotides are detailed in the later chapters along with the corresponding experimental results.

Biotinlyated DNA was anchored on the surface of the coverslips via non-covalent interactions between biotin and streptavidin. Cy3 and Cy5 were attached to a biotinlyated DNA as a FRET donor-acceptor pair, respectively. For doubly-labeled biotinlyated oligonucleotide, Cy3-amidite was directly coupled to the 5' end of the DNA and Cy5-succinimidyl ester was postsynthetically coupled to a C6-amino linker at the 3' end of the DNA. To prevent undesirable G residue quenching effects, a T nucleotide and TTTT overhang were added at 5' and 3' terminus, respectively. The 3' overhang also prevents formation of a dark nonfluorescent state that can form due to the close proximity of the two dyes. For singly-dye-labeled biotinlyated oligonucleotide, only Cy3-amidite was coupled to the 5' end of the DNA.

All the immobilized oligonucleotides have secondary structures with a hairpin-like shape. For immobilization purposes, a biotin was internally attached to DNA hairpins via a dT-biotin phosphoramidite reagent at T28 in the hairpin loop region.

Occasionally, a biotin attachment was alternatively attached to the 3’ terminal site to avoid the perturbation due to the immobilization (see Scheme 2.4 for two types of immobilization).



A. 5’ Cy3 amidite modified oligonucleotide

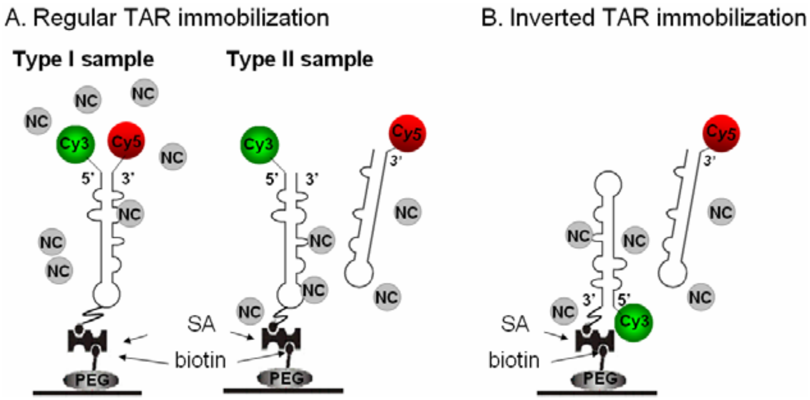
B. 3’ Cy5 succinimidyl ester modified oligonucleotide

C. Biotin-dT phosphoramidite oligonucleotide

Scheme 2.2: Functionalized oligonucleotides.

2.1.2.2 DNA Immobilization Scheme

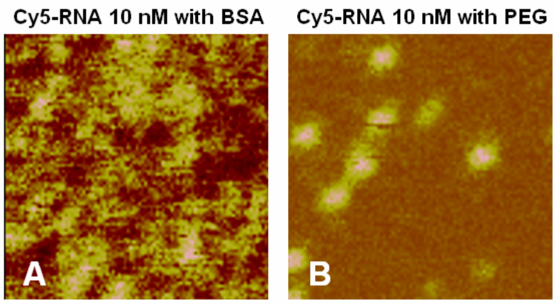
Biotinylated DNA was anchored using two modified surfaces, biotinylated bovine serum albumin (BSA) and biotinylated PEG surface. Prior to the injection of DNA solutions (10–50 pM in buffer A containing 25 mM HEPES, pH 7.3, 40mMNaCl) through the ports into the flow cell, a streptavidin (SA) solution (Molecular Probes, Eugene, OR; 0.2 mg/mL in buffer A) was incubated for 10 min in the cells and washed away. Interactions between DNA and BAS were found to significantly perturb the dynamics of DNA hairpins. This effect was neglected by increasing the concentration of $MgCl_2$ in the buffer. (see Section 3.3.2 for the detailed information).



Scheme 2.3: Immobilization scheme applied for the FRET measurements.

Reannealing of DNA hairpins was performed by incubation at both 80^0C and 60^0C for 2.5 min. A small amount of $MgCl_2$ was added to the DNA solutions before DNA was incubated at 0^0C for 5 min to ensure that the DNA hairpin was at its most stable closed form. The Cy3/Cy5-labeled DNA hairpin (25–50 pM in buffer A and 10 mM $MgCl_2$) was incubated in the flow cell subsequently for 20 min (see scheme 2.3A type I sample) for dynamics measurements of the DNA structure. To produce a sample at single-molecule levels, the surface density of immobilized biotinylated DNA molecules was adjusted by using very dilute DNA solutions (~10 pM) so that individual DNA molecules were well separated from one another. The immobilization allowed one molecule to be monitored over long periods at a time (i.e. > 5 sec). The flow cell was then filled with water and assembled with tubing and a syringe assembly (see Section 2.1.3).

For kinetics measurements of oligonucleotide annealing, Cy3-labeled oligonucleotides were anchored on the surface and complementary Cy5-labeled oligonucleotides were flowed into the cell (see scheme 2.3A type II sample). However, strong Cy5 adsorption onto the BSA surface was observed and resulted in erroneous fluorescence signals (see Scheme 2.4A). To minimize the adsorption of RNA /DNA and proteins onto the surface, we then substituted the biotinylated BSA surface with the biotinylated PEG surface and incubated the cell with SA and biotinylated DNA subsequently.

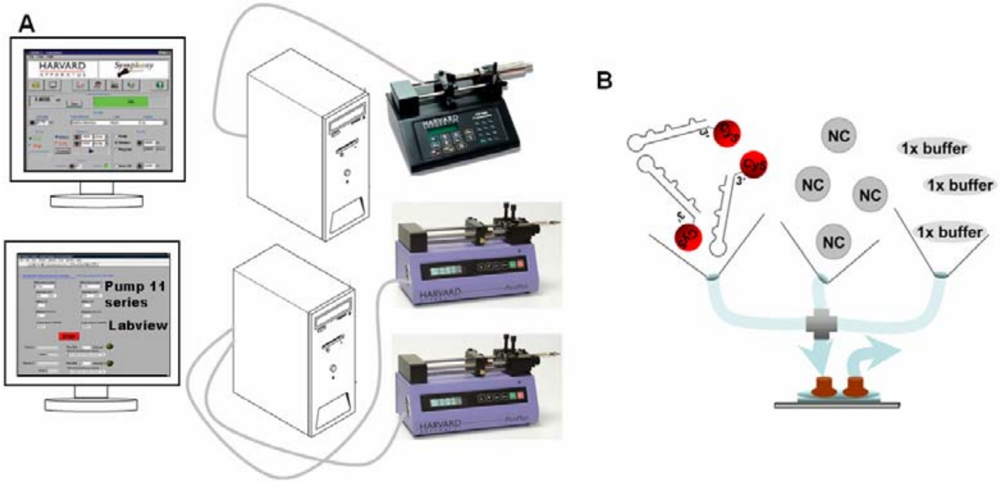


Scheme 2.4: Cy5 adsorption onto the BSA surface but not onto the PEG surface.

The annealing of two oligonucleotides through the hairpin loop-loop interactions were performed using the following immobilization scheme. Cy3-labeled DNA hairpins were inverted and anchored on the surface by switching the biotin linkage from T28 to the 3' terminus (see scheme 2.3B). Thus, perturbation of annealing at the hairpin loop due to immobilization was eliminated.

2.1.3 Micro-fluidic Flow System

After the immobilized DNA samples were prepared (immobilized DNA sample preparation is detailed in section 2.1.2), Teflon tubing was attached to the inlet and outlet port of the flow cell. Three syringes containing three different solutions were connected through tubing and introduced into a 4-way connector and were placed in three syringe pumps (two with the 11 Pico Plus model and one with the PHD 2000 infusion model, Harvard Apparatus, Holliston, MA). (see Scheme 2.5A).



Scheme 2.5: The remote-controlled micro-fluidic flow system.
(A) The remote-controlled pump system and (B) the micro-fluidic flow system are shown.

The syringe pumps were connected to the computer through RS232 serial cables as shown in Scheme 2.5A. Injection of solutions was remote-controlled by two programs, Symphony (Harvard Apparatus) for the PhD 2000 and a Labview program (National Instrument, Austin, TX) for the 11 Pico Plus. The three solutions were introduced into the 4-way connector and generated turbulent mixing then were delivered through tubing into the flow cell. A relatively high flow rate (10 μL/min) was used to ensure thorough and rapid mixing.

Single-molecule measurements were performed at room temperature in buffer A, 0.2~10 mM $MgCl_2$, and an oxygen scavenger system. The oxygen system consisted of a reducing agent, 2-mercaptoethanol 1% v/v (Sigma-Aldrich, St. Louis, MO) and a glucose oxygen scavenger system (*ß*-D(+)glucose 3% w/v (Sigma-Aldrich), glucose oxidase 0.1 mg/mL (Roche Applied Science, Hague Road, IN), and catalase 0.02 mg/mL (Roche Applied Science)) for reducing photobleaching events[96]. The mechanism of the glucose oxygen removal agents is as follows:

Glucose oxidase (GOD) bound to *ß*-D(+)glucose catalyzes the oxidation of *ß*-D(+)glucose into D-Glucono-1,5-lactone which then hydrolyzes to gluconic acid. In order for a redox reaction to occur, oxygen is required to reduce into hydrogen peroxide.

$$glucose + O_2 \xrightarrow{GOD} gluconic\ \ acid + H_2O_2$$

Hydrogen peroxide is then decomposed to water and oxygen as follows:

$$H_2O_2 \xrightarrow{catalase} H_2O + \frac{1}{2}O_2$$

In total, each glucose molecule consumes a half oxygen molecule.

The oxygen scavenger solution should be used with caution. Since glucose reduces into gluconic acid thereby producing hydrogen ions, exposure of glucose solutions to excess oxygen results in a reduction of pH of the buffer A (i.e. a change from pH7.3 to pH5). To maintain solutions at pH7.3, all the solutions were freshly made, immediately transferred to syringes and sealed with plugs (to avoid contact with the air) prior to the experiments.

2.1.4 Nucleocapsid Protein

Wild-type HIV-1 NC (provided by our collaborator, Dr. Musier-Forsyth's group from Ohio State University) was prepared two ways, solid-phase synthesis and expression in *Escherichia coli*. Recombinant NC was made by Dr. Robert Gorelick's group from the National Cancer Institute (for the detailed procedure see Lee et. al, 1998[15]). Synthetic NC proteins were made by Dr. Daniel G. Mullen (from Univeristy of Minnesota). Both proteins were purified by Ms. Brandie J. Kovaleski (from Univeristy of Minnesota).

Synthetic NC protein, with Met 46 mutated to norleucine to avoid oxidation to Met(O), was prepared chemically via Fmoc solid-phase synthesis carried out with the aid of a 433A Peptide Synthesizer (Applied Biosystems, Framingham MA). The C-terminal residue, Asn, was anchored to PAL-PEG-PS resin via its side-chain (i.e. by coupling Fmoc-Asp-*Ot*Bu, as first residue). Deprotection/coupling cycles followed manufacturer recommendations, with all residues double coupled. The fully assembled peptide-resin was cleaved with reagent K [97]. The crude product was purified by reversed-phase HPLC to give a material of ~ 87% homogeneity (as judged by analytical HPLC) (MALDI-MS; $[M + H]_{calc.}$: 6330.2, $[M + H]_{found}$: 6330.1). After lyophilization, the purified synthetic NC was reconstituted by assuming that the recovered white solid contained 100% peptide content. The solid was dissolved (0.6 mL/mg peptide) in reconstitution buffer (40 mM HEPES, 5 mM DTT, 0.1 mM tris(2-carboxyethyl)phosphine hydrochloride) that contained 3.0 equivalents of $ZnCl_2$, and lyophilized for long-term storage. It should be noted that the purified NC before reconstitution often contains a disulfide bond, based on mass spectrometric analysis. The reconstitution buffer contains DTT that reduces the disulfides and allows for correct folding of the Zn fingers to occur. The chaperone activity of the synthetic NC was shown to be higher than NC prepared from *E. coli* expression.

2.2 CONFOCAL SCANNING MICROSCOPY

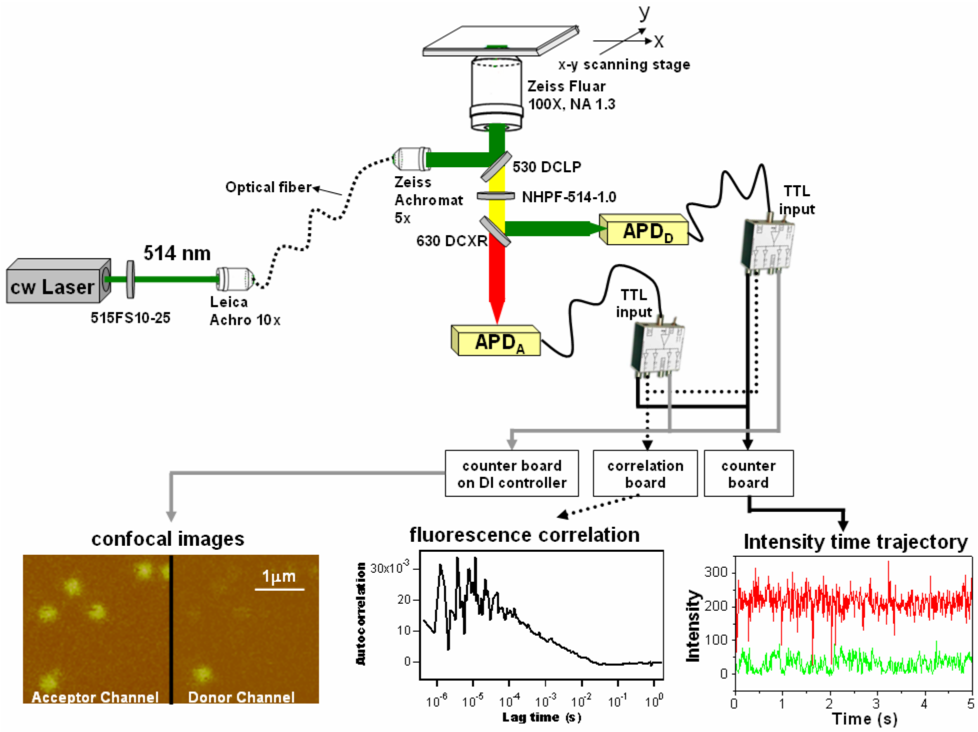
2.2.1 One-Color Continuous Laser Excitation Microscopy

Scheme 2.6 depicts the apparatus for one color continuous excitation microscopy. The detailed description is as follows:

A closed-loop sample scanning stage (NPS-XY-100A, Queensgate, Torquay, Devon, U.K.) was used for imaging and sample positioning. Continuous wave excitation (514 nm, 5 to 10 $\mu W/\mu m^2$) from an argon ion laser (model Reliant 150m, Laser Physics, Inc., West Jordan, UT or tunable ion Laser 543 series, Melles Griot, Carlsbad, CA) was circularly polarized by passing through a polarizer (Melles Griot) and a ¼ wave plate (Newport). The circularly polarized beam was then introduced via an optical fiber (single-mode, operating wavelength at 514/488 nm, Newport Corporation, Irvin, CA), coupled out of the fiber and magnified by a low magnification objective (Zeiss Achromat, 5x) to maximally fill the back aperture of the microscope objective in order to generate a diffraction-limited-spot. The excitation light was filtered by a bandpass filter (515FS10-25, Andover Corporation, Salem, NH) to block scatted laser light, collimated, directed by a dichroic beamsplitter (530 DCLP, Chroma Technology Corp, Rockingham, VT) and focused by a high numerical aperture, oil-immersion, microscope objective (Zeiss Fluar, 100X, NA 1.3) (Carl Zeiss, Oberkochen, Germany) into the sample.

Fluorescence signals were collected by the objective and filtered from scattered excitation light with a holographic Raman notch filter (HNPF-514-1.0, Kaiser Optical Systems Inc., Ann Arbor, MI). Fluorescence signals were separated with a second beamsplitter (Chroma 630 DCXR) and then focused onto two avalanche photodiode (APD) detectors, APD_A and APD_D (see Scheme 2.6). (Perkin Elmer Optoelectronics SPCM-AQR-15, Vaudreuil, Quebec, Canada) via a tube lens inside the microscope. The TTL output signal from the APDs was distributed by a 1:4 fanout TTL driver (Pulse

Research Lab, Torrance, CA) into two counter boards on a nanoscope III controller (Digital Instruments Inc., Woodbury, NY), an ALV 5000/E (ALV-Laser, Langen, Hessen, Germany) correlation board and a counter board, generating confocal scanning images, fluorescence correlation data and fluorescence time trajectories, respectively. This configuration allowed us to simultaneously obtain intensity time trajectories and donor-acceptor intensity cross correlation for each single molecule.



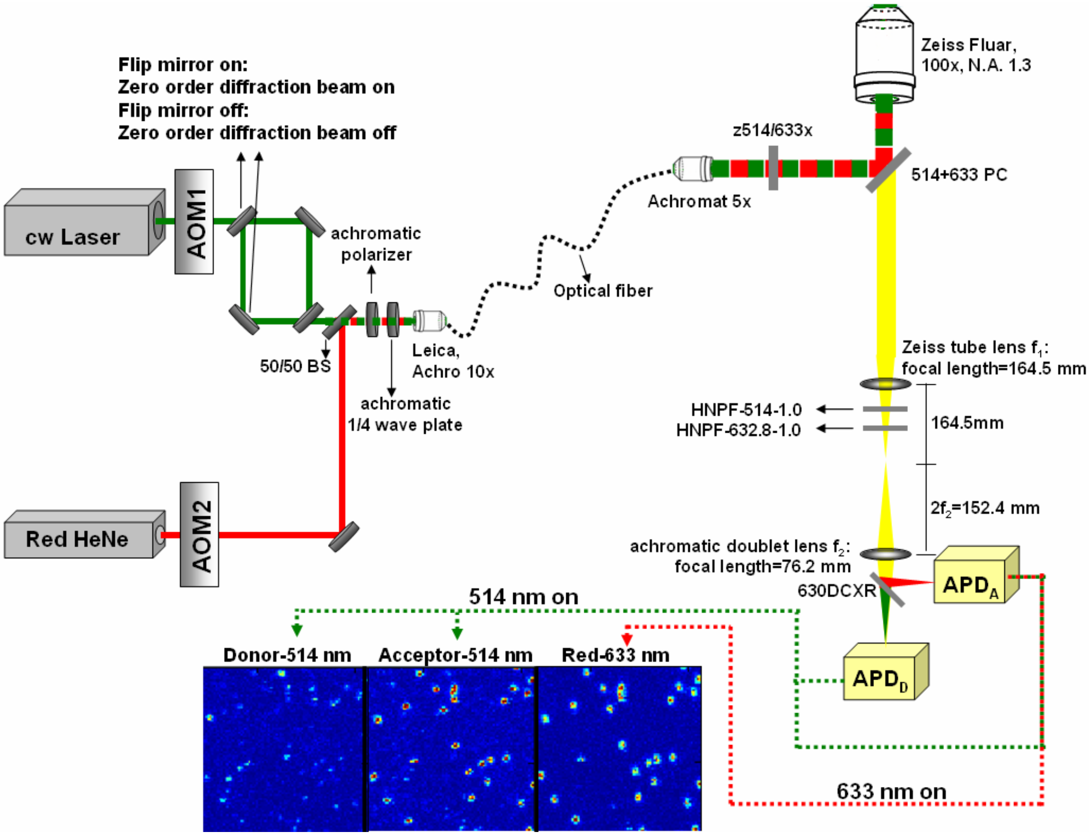
Scheme 2.6: One-color continuous excitation microscopy.

2.2.2 Two-Color Alternating Laser Excitation Microscopy

To achieve the two-color alternating laser excitation fluorescence spectroscopy[98], an additional Red HeNe Laser (633 nm, 3 to 5 μW/μm^2, Melles Griot) was coupled into the same optical fiber in the existing apparatus at a later time (see

Scheme 2.7) using a broadband beamsplitter (50/50 laser splitting, CVI, Vovina, CA). The coupling efficiency for the 514nm laser (30%) is slightly higher than that of 633 nm laser (25%). The laser excitations were pulsed alternately by two acousto-optic modulators (AOMs) (Model 402A-F1, IntraAction Corp., Bellwood, IL) synchronized by a function generator (WaveTek datron 10 MHz DDS, Fluke Corp., Everett, WA) with a 180-degree phase shift. To compensate for chromatic aberration caused by the objective, the 633nm excitation beam was uncollimated and focused via the same microscope objective. The original interference filter was replaced by a customized dual bandpass filter (Chroma z514/633x) to block all other light than 514 and 633 nm. The original dichroic beamsplitter for 514 nm was replaced with a customized dichroic mirror for directing both 514 nm and 633 nm into the objective (Chroma 514+633 PC). An extra holographic Raman notch filter (Kaiser Optical HNPF-632.8-1.0) was added to filter the scattered 633nm excitation light.

Three counters were used to record the photon signals from the APDs. While the 514-nm lasers were on, two counters were used to record the photon signals from APD_A and APD_D due to FRET. While the 633-nm laser was on, an additional counter was used to record the photon signals from APD_A due to the direct excitation of acceptor dyes (referred as the red channel). Acceptor or red images acquired by focusing 514-nm or 633-nm laser beam onto the sample showed that the two beams had no observable difference in the z position of the focal plane (mean photon counts of acceptor images and red images are both optimized using the same focus). The configuration for two-color excitation allowed us to determine if a FRET signal resulted from the presence of an acceptor. Two flip mirrors were added in the 514-nm beam path so either the zero order diffraction beam or the first order diffraction beam could be coupled into the optical fiber (see Scheme 2.7). Therefore, the apparatus can be easily switched between one-color laser continuous excitation and two-color laser alternating excitation.



Scheme 2.7: Two-color alternating laser excitation microscopy.

2.3 SINGLE-MOLECULE FLUORESCENCE RESONANCE ENERGY TRANSFER SPECTROSCOPY

Single-molecule fluorescence intensity time trajectories were recorded using two separate donor and acceptor channels with 1-ms time resolution. The signals $S_i(t)$ were corrected for background ($B_i(t)$) and cross talk (C_i) due to overlapping emission (i.e. the emission from a donor is measured by the acceptor detector) using Equation 2.1:

$$I_i(t) = S_i(t) - B_i(t) - C_i \tag{2.1}$$

where i is donor or acceptor, $I_i(t)$ is the corrected fluorescence intensity of the donor/acceptor dye[99]. The corrected intensity trajectories for the donor and acceptor channels, $I_A(t)$ and $I_D(t)$, respectively, were binned in the appropriate widths (typical with 10-ms binning time (τ_B)). The single-molecule FRET efficiency, $E_{FRET}(t)$ can be defined as:

$$E_{FRET}(t) = \frac{I_A(t)}{I_A(t) + I_D(t)\frac{\phi_A \eta_A}{\phi_D \eta_D}} \tag{2.2}$$

where ϕ_i is the quantum yield of emission of the respective dye, and η_i is the efficiency of each detector. In the experiments, what is measured, however, is the apparent FRET efficiency, $E_A(t)$ defined as:

$$E_A(t) = \frac{I_A(t)}{I_A(t) + I_D(t)} \tag{2.3}$$

In order to convert E_A into E_{FRET}, it is necessary to determine the correction factor $((\Phi_A \eta_A)/(\Phi_D \eta_D))$ for the emission/detection efficiencies. The correction factor was obtained by comparing emission at 100% and 0% (i.e. Cy3-DNA only) energy transfer conditions, and was found to be 1 under the current setup. Thus, the value obtained in each case for E_A represents almost exactly the true value of E_{FRET}.

Blinking events (acceptor reversible photobleaching) were removed prior to data analysis by applying a threshold criterion to the photon counts on the acceptor channel. This criterion filters data points of the acceptor intensity containing only background and cross talk from the donor channel. The threshold criterion equals 2 times the combined uncertainty on the acceptor channel (S_A) resulting from the background (SB_A) and the cross talk (SC_A), as shown in Equation 2.4.

$$2S_A = 2\sqrt{SB_A^{\ 2} + SC_A^{\ 2}} \tag{2.4}$$

Assuming that the only source of uncertainties is photon shot-noise in the background and the cross talk, Equation 4 rearranges to:

$$2S_A = 2\sqrt{< B_A > + < C_A >}$$

where $<B_A>$ is the average background intensity in the acceptor channel and $<C_A>$ is the average cross talk from a donor into the acceptor channel before acceptor photobleaching (Note that for simplicity we have approximated C_A as a constant value; C_A is, however, dependent on the donor intensity, and therefore on hairpin end-to-end distance fluctuations). Under the current experimental setup and conditions, $<C_A>$, the signals of a acceptor channel due to the acceptor emission was $0.25\times<I_D>$ and $<C_D>$ was negligible. Time intervals where acceptor counts were lower to or equal to the threshold value (2 x $(<B_A> + <C_A>)^{(1/2)}$) were removed from the intensity time trajectories. E_A was calculated for each binned point over the entire trajectory. The filtering procedure was not performed for low E_A conditions (i.e. E_A <0.5) due to limited signals.

2.4 FLUORESCENCE CORRELATION SPECTROSCOPY ON IMMOBILIZED MOLECULES

2.4.1 Fluorescence Correlations Calculated from Intensity Time Trajectories

The single-molecule cross correlation curve of $I_D(t)$ vs. $I_A(t)$ ($I_D(t)$ and $I_A(t)$ were from a single-molecule intensity trajectory) was calculated for the following expression:

$$CC(\tau) = \frac{\langle \delta\, I_D(t)\delta\, I_A(t+\tau)\rangle}{\langle I_D\rangle\langle I_A\rangle} = \frac{\langle I_D(t)I_A(t+\tau)\rangle}{\langle I_D\rangle\langle I_A\rangle} - 1 \qquad (2.5)$$

where τ is the lag time. Each single-molecule cross correlation curve was normalized to the autocorrelation of E_A under the assumption that $I_A(t)+I_D(t)$ is constant, that led to:

$$\delta\, I_A(t) = -\delta\, I_D(t)$$

Under this assumption the autocorrelation of E_A (E_AAC)) equaled the autocorrelation of I_A (Equation 6).

$$E_AAC(\tau) = \frac{\langle \delta\, E_A(t)\delta\, E_A(t+\tau)\rangle}{\langle E_A\rangle\langle E_A\rangle} = \frac{\langle \delta\, I_A(t)\delta\, I_A(t+\tau)\rangle}{\langle I_A\rangle\langle I_A\rangle}$$

(2.6)

Replacing $\delta I_D(t)$ by $-\delta I_A(t)$ in Equation 5 we got:

$$CC(\tau) = \frac{\langle -\delta\, I_A(t)\delta\, I_A(t+\tau)\rangle}{\langle I_D\rangle\langle I_A\rangle} \qquad (2.7)$$

Multiplying Equation 7 by ($-<I_D>/<I_A>$) the cross correlation was normalized to the E_AAC, see Equation 8.

$$E_AAC(\tau) = CC(\tau)\left(\frac{-<I_D>}{<I_A>}\right) \qquad (2.8)$$

Single-molecule E_AAC curves obtained under the same condition were superimposed, averaged and fitted with proper exponential functions to extract the kinetic rate of dynamics of DNA hairpins. The E_AAC amplitude at $\tau=0$, however, was removed due to the domination of shot noise.

2.4.2 Real-Time Fluorescence Correlation Measurement

Donor-acceptor cross correlations were also obtained with an ALV 5000/E correlation board with 0.5-μs time resolution. Data was acquired for a total of 4 seconds. The data from any single molecule that photobleached prior to finishing acquisition was discarded.

The real-time correlations were calculated by the ALV program while the correlation board was collecting the signals from a donor and an acceptor. Thus the real-time correlation data contained true signals from donor/acceptor emission as well as signals from background and cross talk. Intensity time trajectories from a donor and acceptor were provided by the correlation board with a lower time resolution (~26 ms), giving the averaged total signals from a donor ($<S_D>$) and an acceptor ($<S_A>$). Thus the donor-acceptor correlations obtained in real-time ($CC_R(\tau)$) could be expressed as Equation 2.9:

$$CC_R(\tau) = \frac{\langle \delta\, S_D(t) \delta\, S_A(t+\tau) \rangle}{\langle S_D \rangle \langle S_A \rangle} = \frac{\langle S_D(t) S_A(t+\tau) \rangle}{\langle S_D \rangle \langle S_A \rangle} - 1 \tag{2.9}$$

To correct for the background and cross talk from the total signals, intensity time trajectories from the background in two channels were also acquired for a total of 4 seconds after both Cy3 and Cy5 were photobleached. This gave an average background $<B_i>$ for both channels. $<I_i>$ could be obtained by subtracting $<B_i>$ and C_i from $<S_i>$ (see Equation 2.1 for the formula). Since $\delta S_i(t)$ is the fluctuation of emission at a specific time relative to the average emission at the whole acquisition time period, $\delta I_i(t)$ approximated $\delta S_i(t)$. Therefore, $C(\tau)$ was estimated using $C_R(\tau)$ by multiplying Equation 9 by $((<S_D>\times<S_A>)/(<I_D>\times<I_A>))$ as Equation 10:

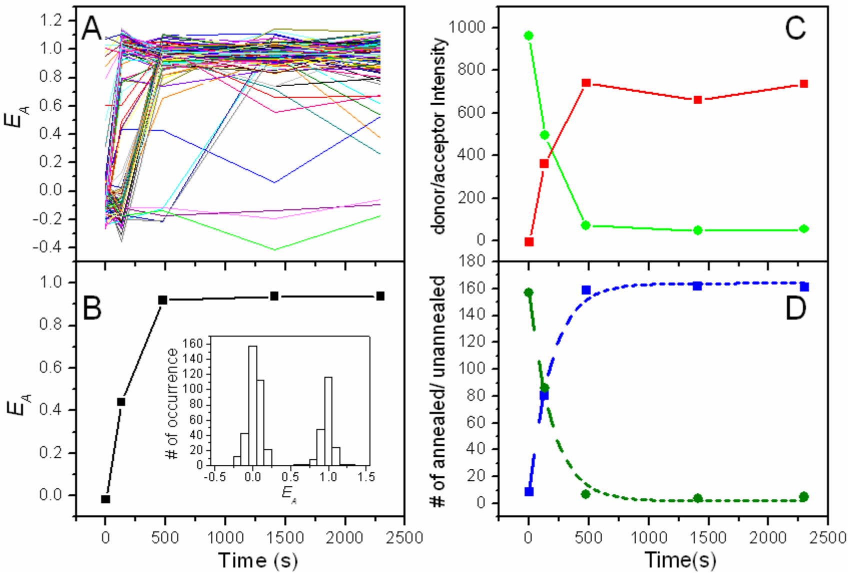
$$CC(\tau) = CC_R(\tau) \frac{\langle S_D \rangle \langle S_A \rangle}{\langle I_D \rangle \langle I_A \rangle} = \frac{\langle \delta\, I_D(t) \delta\, I_A(t+\tau) \rangle}{\langle I_D \rangle \langle I_A \rangle} \tag{2.10}$$

We were then able to estimate E_AAC by multiplying Equation 2.10 by ($-<I_D>/<I_A>$) as previously described (see Equation 2.8 for the detailed expression).

2.5 TIME RESOLVED CONFOCAL SCANNING IMAGING

Single-molecule kinetic experiments of the annealing of immobilized Cy3-oligonucleotides to the reacting Cy5-oligonucleotides in the solutions were accomplished by recording multiple confocal scanning images of a 30 μm x 30 μm region at various times after introduction of the Cy5-labeled species to the flow cell. Each image frame contained 500-600 single molecule spots corresponding to 1 molecule per 3 μm by 3 μm area. The total scan time for each image frame was 2-3 min and donor and acceptor images were collected simultaneously.

Photobleached molecules were discarded from the statistical analysis. The images were analyzed automatically by a Matlab program that determined the location and intensity of each molecule in the donor and acceptor images for each frame. An E_A value was calculated for each molecule using the observed donor/acceptor intensities, and corrected for crosstalk and background intensity. The data analysis yielded E_A histograms, E_A time trajectories for each molecule, $<E_A>$, $<I_i>$ and number of donors/acceptors for each frame.



Scheme 2.8: Results from analyzing multiple confocal images for the SM-FRET measurements.

(A) E_A time trajectories for each molecule (B) $<E_A>$ (C) $<I_i>$ and (D) number of donors (shown in green)/acceptors (shown in red) at 5 acquisition times for each image. (B, Inset) An ensemble E_A histogram shows a well resolved bimodal distribution in which high and low E_A population represents annealed and unannealed species, respectively.

For each molecule in each frame, the $0.5E_A$ value was used to classify each molecule as either the un-annealed reactant or the annealed product (see the inset in scheme 2.8 B). E_A histograms for each frame showed well resolved peaks for annealed and unannealed species (see chapter 4 for the detailed information). The number of the annealed products (Scheme 2.8 D) was then converted to % of annealed molecules. The % of annealed molecules vs. time curves were then fitted to a first-order kinetic model yielding a rate constant k or reaction time $\tau=1/k$.

For two-color alternating laser excitation experiments, the number of Cy5-DNA on an immobilized Cy3-DNA was determined by two separate procedures, direct counting and ensemble averaging. Both procedures require a localized Cy5 emission intensity on an immobilized Cy3-oligonucleotide. This value was determined from the peak in an ensemble histogram of the Cy5 emission intensities on an immobilized Cy3-oligonucleotide acquired from an image obtained with no Cy5 in the solution (i.e. replace Cy5-labeled species in the solution with the buffer). This value was a suitable threshold for counting the number of Cy5 molecules and a suitable mean single-molecule intensity of a Cy5-labeled molecule.

In the direct counting procedure, this threshold was used to determine whether a Cy5-molecule was present at a particular immobilized Cy3-hairpin. However, this procedure failed due to the following two reasons. First, a high concentration of Cy5-labeled oligonucleotides in the solution resulted in a poor signal to noise ratio. Second, large Cy5-species adsorbed onto the surface (Cy5 adsorption was especially problematic for Cy5-labeled RNA) resulted in high Cy5 intensity on specific spots even with no immobilized Cy3-hairpins.

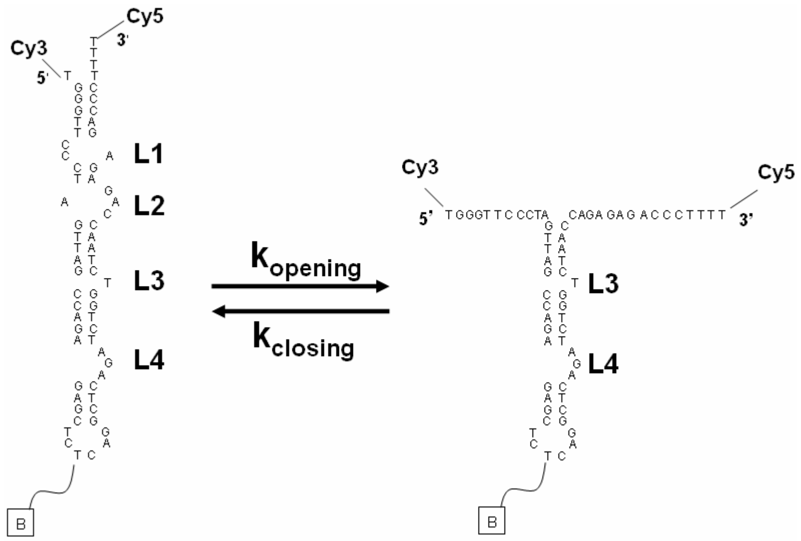
While the direct counting method failed, an ensemble averaging procedure worked well in the presence of Cy5-labeled species up to much higher background levels.

The number of Cy5 molecules on an immobilized single molecule was determined by averaging the Cy5 emission intensity on all the selected Cy3-labeled immobilized molecule spots in each image frame and dividing the mean Cy5 intensity by the previously determined mean intensity for a single Cy5-labeled species. Under the conditions of small background noise (i.e. low Cy5-oligonucleotide concentrations in the solution), the number of Cy5 molecules on an immobilized Cy3-haripin estimated by the two procedures reached agreement. A disadvantage of the ensemble averaging procedure is that it gives the total number but not the location of bound Cy5-oligonucleotides on immobilized Cy3-haripins.

Chapter 3: Secondary Structure and Secondary Dynamics of DNA Hairpins Complexed with HIV-1 NC Protein

3.1. INTRODUCTION

The effect of NC binding on the conformations of TAR DNA hairpins has previously been investigated by ensemble FRET measurements, fluorescence lifetime measurements and fluorescence correlation spectroscopy[68,69,71,100]. These data demonstrate that bound NC shifts the TAR DNA hairpin equilibrium towards "open" conformations. However, the averaging inherent in ensemble FRET measurements makes it difficult to identify these specific " open" conformations [68,69,71,100].



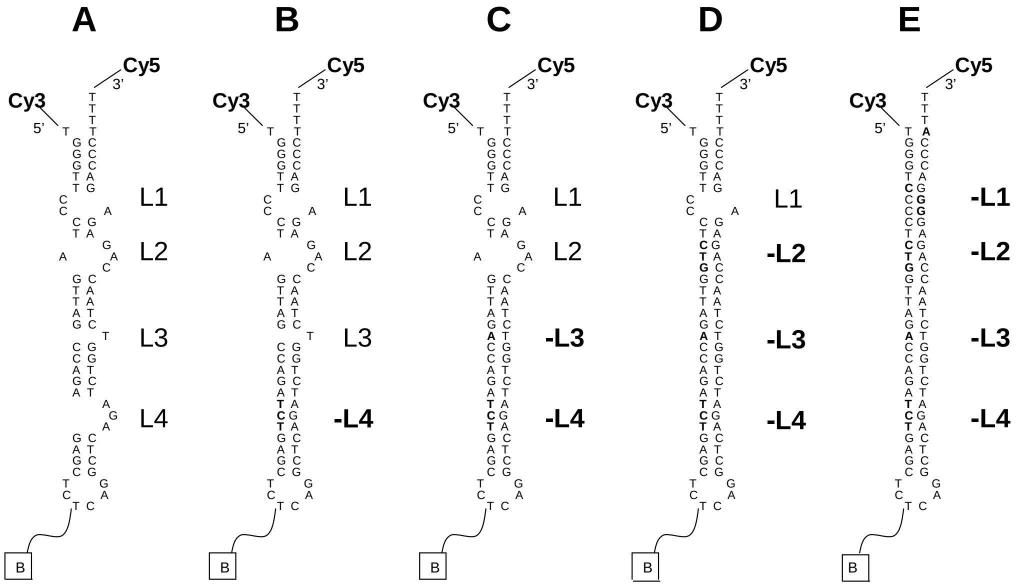
Scheme 3.1: TAR DNA hairpin dynamics with 445nM NC.

In order to obtain more detailed information on the conformations of TAR DNA/NC, the first time-resolved SM-FRET measurements on TAR DNA hairpins and hairpin mutants was performed in the presence and absence of NC protein. The single-molecule results clearly demonstrate that bound NC shifts the equilibrium of the secondary structure of TAR DNA hairpins from a fully "closed" conformation to a specific "partially-open" conformation with the two terminal L1L2 regions "open" or unwound and the other L3L4 regions closed (see the Scheme 3.1). In addition, the data show that the two terminal L1L2 in the TAR DNA/NC complex undergo a rapid opening/closing process. The observed partially-open TAR DNA/NC conformation seems ideally suited for the promotion of NC-catalyzed DNA/RNA annealing, since the open terminal stems of TAR DNA possess 21 unpaired bases accessible for the annealing of incoming TAR RNA.

3.2. IMMOBILIZED DNA AND STANDARD DEVIATION OF E_A HISTOGRAMS

3.2.1 DNA Hairpins Employed in this Study

We recorded the trajectory and the distribution of $E_A(t)$ for single molecules of various donor-acceptor (Cy3-Cy5) labeled WT and mutant TAR DNA hairpin constructs with different numbers of internal bulges/loops (see Scheme 3.2). We determined the number of hairpin stems affected by NC by comparing the data acquired with the different constructs in the presence and absence of NC. Previous work has demonstrated that maximum NC destabilizing activity is obtained under saturating protein concentrations, where one NC protein is bound to eight nucleotides[22,35]. Maximum annealing rates between TAR DNA and TAR RNA hairpins occur for [NC] ≥ 500 nM[101]. Our experiments were performed with saturating NC concentrations (445nM NC) close to this range. All the immobilized DNA samples were prepared using BSA surface (see section 2.1.1.1).



Scheme 3.2: Secondary structures of the oligonucleotides employed in the SM-FRET studies.

The structures were predicted by the mfold program (http://www.bioinfo.rpi.edu/applications/mfold/old/dna/). (A) TAR: a 64-nt DNA hairpin. (B) –L4TAR: a 67-nt TAR DNA mutant for which one internal bulge was deleted. (C) –L3L4TAR: a 68-nt TAR DNA mutant for which two internal bulges were deleted. (D) –L2L3L4TAR: a 71-nt TAR DNA mutant for which three internal bulges were deleted. (E) –L1L2L3L4TAR: a 74-nt TAR DNA mutant for which the four internal bulges were deleted.

3.2.2 Standard Deviation of E_A Distribution

The photobleaching events of donor-acceptor intensity time trajectories were removed by applying a threshold criterion to the photon counts in the acceptor channel as previously described in Section 2.3 (also see Figure 3.1). Figure 3.3 top shows that the donor and acceptor fluorescent intensity (shown in green and in red, respectively) acquired with original time resolution, $\tau_B = 1$ ms. The dotted black line corresponds to the threshold value given by $(2 \times (<B_A> + <C_A>)^{(1/2)})$ (see section 2.3 for the detailed information). The ratio of acceptor to the sum of acceptor plus donor intensities was used to calculate the trajectory of $E_A(t)$ shown below in Figure 3.3.

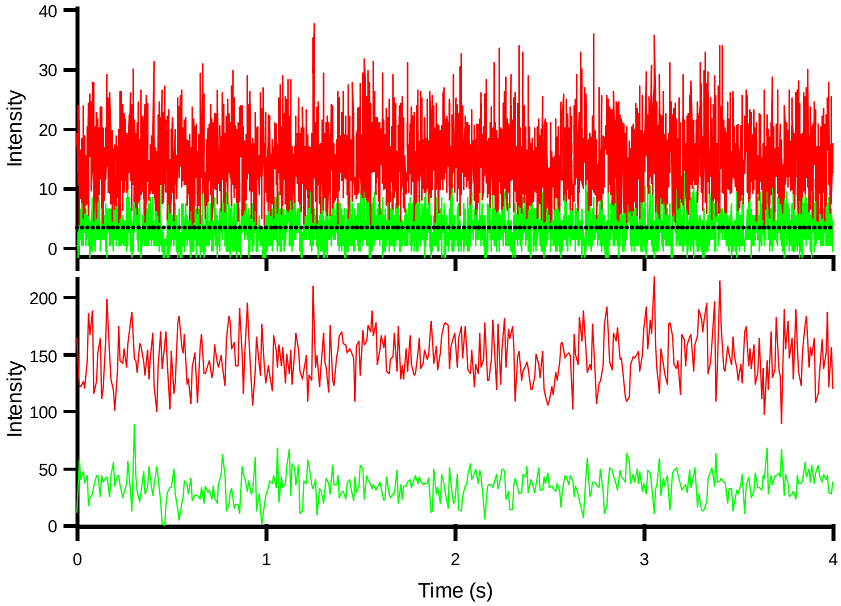


Figure 3.1: Intensity time trajectories for the representative –L3L4TAR single molecule in the presence of NC, buffer A and 0.2 mM $MgCl_2$.

An analytical expression for the experimental $E_A(t)$ distribution standard deviation (σ_{E_A}) was derived using error propagation, assuming that the only source of uncertainties is photon shot-noise in the acceptor and donor intensities[99] and in the acceptor and donor backgrounds (Equation 3.1). We calculated the standard deviation for any value of E_A using the minimum experimental total fluorescence intensity counts detected, 140 counts and the counts from donor and acceptor background, 10 and 20 counts, respectively (both with τ_B =10 ms).

$$\left(\sigma_{noise}\right)=\sqrt{\frac{\left(\langle I_D\rangle-\langle B_D\rangle\right)^2\times\left(\langle I_A\rangle+\langle B_A\rangle\right)}{\left(\langle I_A\rangle-\langle B_A\rangle+\langle I_D\rangle-\langle B_D\rangle\right)^4}+\frac{\left(\langle I_A\rangle-\langle B_A\rangle\right)^2\times\left(\langle I_D\rangle+\langle B_D\rangle\right)}{\left(\langle I_A\rangle-\langle B_A\rangle+\langle I_D\rangle-\langle B_D\rangle\right)^4}}$$

(3.1)

Figure 3.2 illustrates the dependence of σ_{E_A} with E_A. The values were employed to determine the confidence intervals in experimental $E_A(t)$ distributions.

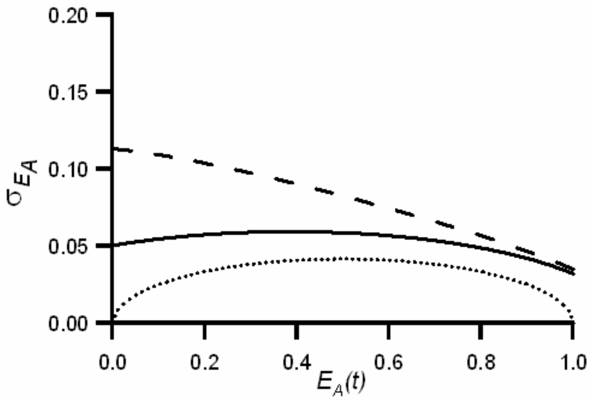


Figure 3.2: Theoretical σ_{E_A} of $E_A(t)$ distribution.

(Solid line): Curve obtained applying Equation 3.1 and the following experimental parameters: total fluorescence intensity counts 140 using τ_B = 10 ms, donor background 10 counts per bin, acceptor background 20 counts per bin. (Dashed line): Theoretical σ_{E_A} calculated with the same parameters and for a 25% donor cross talk into the acceptor channel. (Dotted line): Theoretical σ_{E_A} of $E_A(t)$ distribution calculated for a total fluorescence intensity of 140 counts and no background[99].

3.3. RESULTS AND DISCUSSION

3.3.1 HIV-1 NC Effect on the Distribution of Hairpin End-to-end Distances

Any single molecule is representative of the ensemble under our experimental conditions, i.e., the system is ergodic. Individual single molecule trajectories and distributions of $E_A(t)$ are indistinguishable from each other when recorded under the same experimental conditions. The mean E_A values measured for individual single molecules ($<E_A>_{mol}$) are also very similar to each other (Figure 3.3, column 1).

Hairpins with two or more internal bulges are predominantly found with the two terminal stems open at equilibrium with 445 nM NC. An $<E_A>_{mol}$ ~0.77, σ_{E_A} = 0.07 was measured for TAR DNA single molecules with NC flowed at a rate of 2 µl/min (Figure 3.3 A). The same values were obtained for mutant forms of TAR DNA hairpins

(-L4TAR, see Scheme 3.2 B) for which the internal bulge L4 and both internal bulges L3 and L4 (-L3L4TAR, see Scheme 3.2 C) were deleted. Inspection of the single molecule trajectories of $E_A(t)$ reveals that the hairpins sporadically closed to the conformation with E_A~ 1 (see right panels in Figure 3.3 A and B). The closing events are also evident from the unsymmetrical distribution of $E_A(t)$ (see Figures 3.3 A and B). From the distribution of $E_A(t)$, as much as 10% of the events can be assigned to a closed hairpin conformation.

NC induces larger donor-acceptor dye separations in a structure with two internal bulges compared to one internal bulge. Thus the distribution of $E_A(t)$ for a third TAR DNA mutant form conserving only the terminal internal loop L1 missing (–L2L3L4TAR, Scheme 3.1 D) has a mean value $<E_A>_{mol} = 0.90$, $\sigma_{E_A} = 0.05$ with 445 nM NC (see Figure 3.3 C).

A TAR DNA mutant for which all four internal loops were deleted (-L1L2L3L4TAR, Scheme 3.2 E) is characterized by a fully closed conformation with or without NC, $<E_A>_{\mathrm{mol}}$ ~0.97, $\sigma_{E_A} = 0.04$ (Figures 3.3 D and 3.3 F, respectively). In fact, the individual $E_A(t)$ trajectories for –L1L2L3L4TAR DNA with or without NC are indistinguishable from trajectories for –L3L4TAR hairpins without NC (Figure 3.3 E). This result confirms previous reports on the critical importance of internal bulges in the hairpin structure to observe an NC effect[69].

The data provide direct evidence at the individual hairpin level that NC activity in TAR DNA is limited to internal loops L1 and L2 and does not involve the internal bulges L3 and L4 (see scheme 3.1 *left* for L1-L4 designation). Thus, all hairpin constructs conserving the two internal loops L1 and L2 have identical distributions of $E_A(t)$ when complexed with NC, irrespective of the presence of other destabilizing internal bulges.

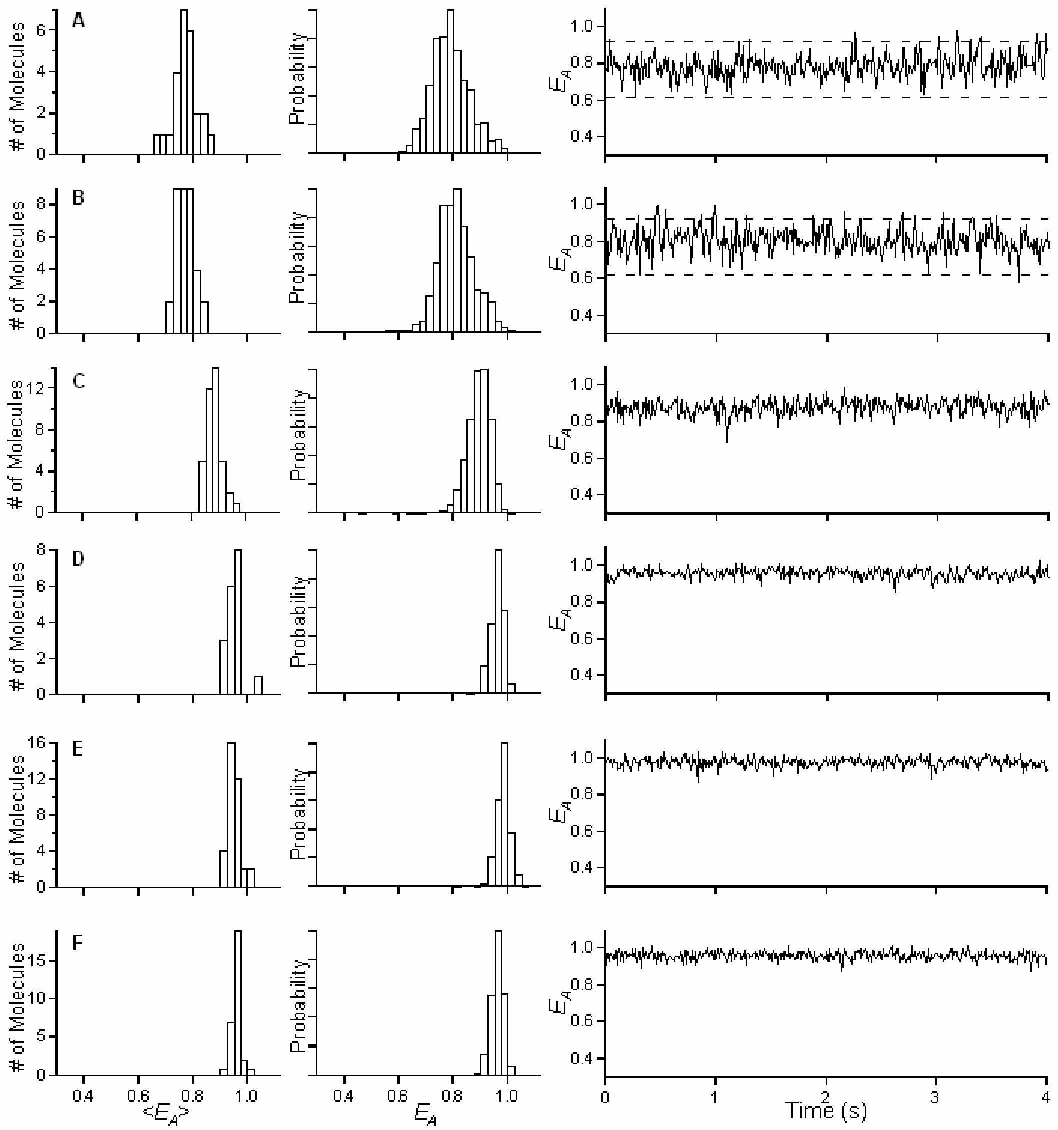


Figure 3.3: $<E_A>_{mol}$ and E_A histograms and E_A time trajectories of various WT and mutant TAR DNA hairpins.

Distribution of $<E_A>_{mol}$ values is shown in the first column. Distribution and trajectory of $E_A(t)$ for a representative single molecule (τ_B= 10 ms) is shown in the second and third columns. Rows A, B, C and D correspond to data acquired with 445 nM NC and 0.2 mM $MgCl_2$. (A) TAR DNA single molecules. (B) –L3L4TAR DNA single molecules. (C) –L2L3L4DNA TAR single molecules. (D) –L1L2L3L4TAR DNA single molecules. Also shown are the corresponding figures for (E) – L3L4TAR DNA and (F) –L1L2L3L4TAR DNA with no NC and with 0.2 mM $MgCl_2$. Overlaid on the trajectories of TAR DNA/NC and – L3L4TAR/NC (panels A and B) is shown the 99% confidence interval for a distribution centered at E_A = 0.77. Any point outside the range ($\pm 2.6 \times \sigma_{E_A=0.77}$) does not correspond to an open conformation, where $\sigma_{E_A=0.77}$ is the standard deviation estimated from shot noise for E_A=0.77.

In comparison, the hairpin conserving only the last internal loop can only access a conformation with a smaller end-to-end distance when complexed with NC. Finally, the construct where all the internal loops and bulges have been deleted is only found in the closed conformation.

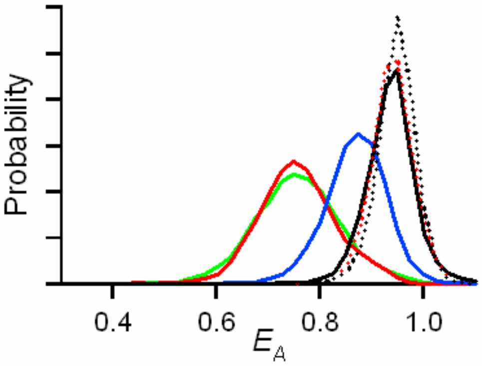


Figure 3.4: Normalized ensemble distribution of $E_A(t)$ of various WT and mutant TAR DNA.

The histograms is constructed by combining the individual single molecule distribution of $E_A(t)$ from 20-30 molecules acquired with τ_B = 10 ms. Curves correspond to TAR DNA (green), –L3L4TAR (red), –L2L3L4TAR (blue), and –L1L2L3L4TAR (black). All measurements were carried out in the presence of 445 nM NC and 0.2 mM $MgCl_2$. Also shown are the normalized ensemble distribution of $E_A(t)$ for –L3L4TAR (dashed red line) and –L1L2L3L4TAR (dashed black line) with no NC and with 0.2 mM $MgCl_2$.

Approximately 20% of the single molecules observed for TAR DNA/NC (but not any of the mutant hairpin structures) exhibited extremely slow FRET dynamics (relaxation lifetimes >>1 second) with slowly varying E_A amplitude changes. We assign the single molecules with slow dynamics to hairpins that are perturbed by binding to the BSA surface. This is supported by the following observations: (1) analogous slow FRET dynamics are not observed in bulk ensemble FRET measurements on non-immobilized TAR/DNA; (2) a slight increase of the $MgCl_2$ concentration in the buffer suppresses the formation of slow FRET dynamics, due presumably to better hairpin/surface screening; and 3) FRET dynamics on the many second time scale is too slow to be associated with DNA hairpin opening/closing reactions.

Care has been taken to rule out any influence that fluorophore blinking (reversible photobleaching) or fluorophore quenching by NC might have on the observed trajectories and distributions of $E_A(t)$ (see Figure 3.1). Thus, we note that no emission quenching is observed upon NC addition to the construct that lacks internal loops and bulges. On the other hand, Cy5 blinking events were removed from the $E_A(t)$ trajectories as described in the experimental section. We conclude that the $<E_A>_{\mathrm{mol}} < 0.97$ observed in hairpins with internal loops/bulges (Figures 3.3 *A*-*C*) in equilibrium with NC is a direct result of NC-hairpin interactions.

Consistent with a closed hairpin structure, the distribution of $E_A(t)$ values for –L3L4TAR single molecules in the absence of NC has a mean value $<E_A>_{mol} = 0.97$, $\sigma_{E_A} = 0.04$, (Figure 3.3 *E*). Figure 3.4 portrays the normalized ensemble distributions of $E_A(t)$ for the different hairpins with and without NC to facilitate their comparison.

3.3.2 HIV-1 NC Effect on Dynamics of Hairpin End-to-end Distances

Hairpin opening-closing relaxation rate constants k_r were obtained from cross correlation analyses of the single-molecule donor-acceptor intensity time trajectories. Briefly, in this analysis fluorescence intensity fluctuations in the donor channel are correlated with fluorescence intensity fluctuations in the acceptor channel to quantify the temporal evolution of the system in equilibrium[81]. From the cross correlation analyses we were able to assign the rate constants associated with the observed hairpin conformational distributions.

The correlation analysis of the data reveals dynamics in the milliseconds time domain for hairpins with one or more internal bulges in equilibrium with NC.

Consistent with dynamics reflecting FRET efficiency fluctuations, donor and acceptor intensities are anticorrelated i.e., an increase in donor intensity is accompanied by a decrease in acceptor intensity[84]. *CC* (Cross correlation) was converted to E_AAC (autocorrelation of E_A) as described in the Section 2.4.

Table 3.1: E_AAC relaxation rate constants and amplitudes for various hairpins in the presence of 445 nM NC and with no NC

Hairpin Construct	$^*k_r$ (s^{-1})	$^*A$	$^\dagger k_r1$ (s^{-1})	$^\dagger A_1$	$^\dagger k_r2$ (s^{-1})	$^\dagger A_2$
TAR DNA/NC	3×10^2	6×10^{-3}	2×10^2	8×10^{-3}	3×10^3	10×10^{-3}
-L3L4TAR/NC	2×10^2	6×10^{-3}	2×10^2	8×10^{-3}	2×10^3	9×10^{-3}
‡-L3L4TAR/NC	0.7×10^2	2×10^{-3}	0.8×10^2	4×10^{-3}	0.6×10^3	3×10^{-3}
-L2L3L4TAR/NC	4×10^2	2×10^{-3}	5×10^2	6×10^{-3}		
-L1L2L3L4TAR/NC	§N.A.	§0	§N.A.	§0		
-L3L4TAR	§N.A.	§0	§N.A.	§0		
-L1L2L3L4TAR	§N.A.	§0	§N.A.	§0		

*Obtained from the analysis of $I_A(t)$ and $I_D(t)$ trajectories (1 ms to 4 s range). †Simultaneously measured with an ALV 5000/E correlation board (0.5 μs to 4 s range). ‡Measured in the presence of 2.5 mM $MgCl_2$. *CC* was converted to E_AAC as described in the Section 2.4. The correlations of individual single molecules were averaged to obtain the reported ensemble value. The error in the reported values (± 50%) was determined from the standard deviation of the distributions of k_r and A for individual single molecules. §N.A. indicates not assigned. Amplitudes $\leq 0.5 \times 10^{-3}$ are indistinguishable from noise, therefore the relaxation rate constant values could not be assigned.

Table 3.1 lists the relaxation rate constants and amplitudes derived from the cross correlation analysis of $I_A(t)$ and $I_D(t)$ trajectories for TAR DNA/NC, -L3L4TAR/NC and –L2L3L4TAR/NC (see columns 1 and 2). A single exponential relaxation with a rate constant $k_r = 4 \times 10^2\ s^{-1}$ was measured for –L2L3L4TAR/NC. The single exponential decay indicates that this is a two-state transition between an open and a closed hairpin conformation. The E_A autocorrelation relaxation rate constants for TAR DNA/NC and for –L3L4TAR/NC have values of k_r = 3 x 10^2 s^{-1} and 2 x 10^2 s^{-1}, respectively. Conformational equilibration (relaxation) occurs on a timescale of a few milliseconds, much shorter than the observation time (~ 4-20 sec) for each hairpin, ensuring equilibrium sampling in the observed single-molecule data.

The relaxation rate constant in a two state system (open-closed) depends linearly on both opening and closing rate constants ($k_r = k_{opening} + k_{closing}$)[102]. The predominantly open hairpin conformation directly observed from the distribution of $E_A(t)$ reveals that for TAR DNA/NC or –L3L4TAR/NC the opening rate constant is much larger than the closing rate constant. As much as 10% of the E_A events can be assigned to a closed hairpin conformation. In a two state model this would indicate that opening and closing rate constants for TAR DNA/NC or –L3L4LTAR/NC are in the range $k_{closing} \leq 3 \times 10^1\ s^{-1}$ and $k_{opening} \geq 2.5 \times 10^2\ s^{-1}$.

In an attempt to shift the equilibrium from the predominantly open hairpin conformation directly observed from the distribution of $E_A(t)$ for TAR DNA/NC and –L3L4TAR/NC we increased the $MgCl_2$ concentration by ~10-fold. Figure 3.5 illustrates the distribution of $<E_A>_{mol}$ values (left) and a representative single-molecule $E_A(t)$ distribution (center) and time trajectory (right) in the presence of 2.5 mM $MgCl_2$. Consistent with a stabilization of the closed hairpin by $MgCl_2$[103] the equilibrium was shifted to the closed conformation following addition of 2.5 mM $MgCl_2$. Under these conditions the measured relaxation rate constant is ~ 0.7 x 10^2 s^{-1} and the closed-open equilibrium constant is $K_{closed\text{-}open}$ ~ 1. If a two state model applies to these conditions opening and closing rate constants for –L3L4TAR/NC are $k_{closing} \sim k_{opening} = 0.4 \times 10^2\ s^{-1}$. Two distinct peaks at ~0.8 and 0.97 were resolved in the distribution of $E_A(t)$. These observations support the conclusion that the dynamics of hairpin opening-closing relaxation in the presence of NC occur in the millisecond time domain.

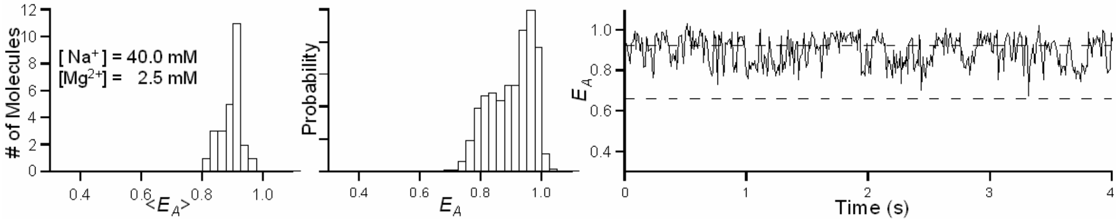


Figure 3.5: $<E_A>_{mol}$ and E_A histograms and $E_A(t)$ for a representative –L3L4TAR hairpin.

Distribution of $<E_A>_{mol}$ values is shown in the left panel. Distribution and trajectory of $E_A(t)$ is shown in the middle and right panel. Data acquired with –L3L4TAR/NC with 445 nM NC and 2.5 mM $MgCl_2$ using $\tau_B = 10$ ms.

The observation of large amplitude NC-induced DNA conformational fluctuations on the millisecond timescale reported here is in contrast to previous work reporting NC induced DNA conformational fluctuations on the microsecond timescale[68,69]. The latter study was carried out using FCS, a technique that is unable to measure slow dynamics, using a related cTAR DNA and truncated NC system.

In order to determine if faster events escaped our detection, we cross correlated the donor and acceptor intensities of single immobilized hairpins over a time range of 0.5 μs to 4 seconds with an ALV5000/E correlator board. Figure 3.6 portrays the ensemble E_AAC derived from donor-acceptor cross correlations. No fast components are observed in –L2L3L4TAR/NC for which the relaxation rate is similar to that obtained from the $E_A(t)$ trajectories (Table 3.1, compare columns 2 and 4). The relaxation for TAR DNA/NC and –L3L4TAR/NC spanned from microseconds to milliseconds with approximately equal weights in both time domains. A biexponential fit to the data generates relaxation rate constants of 2×10^2 s^{-1} and 3×10^3 s^{-1} for TAR DNA/NC and 2×10^2 s^{-1} and 2×10^3 s^{-1} for –L3L4TAR/NC, respectively (Table 1, columns 4 and 6). In the presence of 2.5 mM $MgCl_2$, the decay for –L3L4TAR/NC is biexponential with relaxation rate constants of 0.8×10^2 s^{-1} and 6×10^2 s^{-1}.

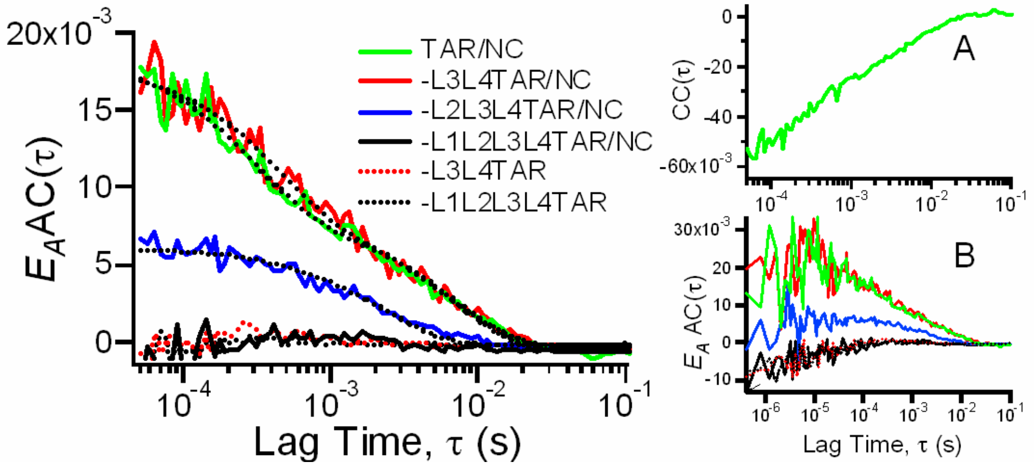


Figure 3.6: Ensemble E_AAC (E_A autocorrelation) and fit to the experimental data from various WT and mutant DNA hairpins with/without NC.

Each ensemble was constructed by superimposing the single-molecule E_AAC from at least 18 immobilized hairpins of TAR DNA (green), –L3L4TAR (red), –L2L3L4TAR (blue), and –L1L2L3L4TAR (black) in the presence of 445 nM NC and 0.2 mM $MgCl_2$. Also shown are the ensemble E_AAC for the –L3L4TAR (dashed red line) and –L1L2L3L4TAR (dashed black line) with no NC and 0.2 mM $MgCl_2$. The E_AAC curves at times $< 5 \times 10^{-5}$ s have been omitted from this figure since they are dominated by Cy5 photophysical events which are unrelated to the FRET behavior. Insets: (A) The corresponding ensemble *CC* for TAR DNA. (B) The full range E_AAC curves from are shown from 5×10^{-7} s to 0.1 s.

Although we have not been able to assign the TAR DNA hairpin conformational distribution associated with the fast relaxation dynamics observed in the correlation analysis ($k_r = 3 \times 10^3$ s^{-1}), it is not surprising that such a complex system involving the interaction of many NC proteins with the TAR DNA hairpin exhibits multiexponential opening-closing relaxations.

We performed control experiments to determine the effect, if any, that Cy5 photophysical properties have on the correlation curves measured[104,105]. Experiments were done with the –L3L4TAR in the absence of NC and with the –L1L2L3L4TAR in the absence and presence of NC. The cross correlation for the controls had an amplitude ~ 0 after 2 × 10^{-4} s. There were no donor-acceptor anticorrelated fluctuations with lifetimes of 100 microseconds or longer. Cross correlations at times < 2 x 10^{-4} s appear to be dominated by Cy5 cis-trans isomerization, unrelated to FRET dynamics.

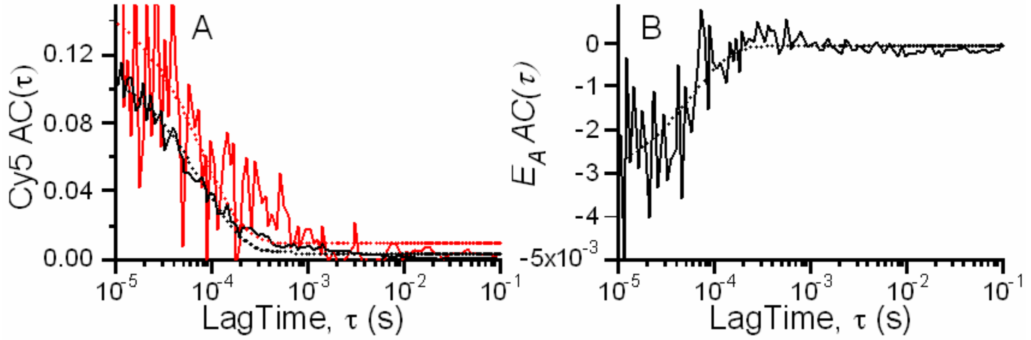


Figure 3.7: Ensemble Cy5 *AC(τ)* and $E_A AC(\tau)$ and fit to the experimental data obtained with –L1L2L3L4TAR DNA hairpins.

(A) Cy5 intensity ensemble *AC(τ)* and fit to the experimental data determined from ~ 20 individual donor-acceptor labeled –L1L2L3L4TAR. Black: in the presence of buffer A and an oxygen scavenger system, the excitation power was ~5 to 10 μW/μm^2. Red: determined in an air equilibrated buffer solution containing glucose 3% w/v; the excitation power was ~1.5 to 3 μW/μm^2. (B) Ensemble $E_A AC(\tau)$ (solid line) and the fitting curve (dot line) to the experimental data. The ensemble was obtained by superimposing the single-molecule $E_A AC(\tau)$ of donor-acceptor labeled –L1L2L3L4TAR. In all cases the donor was directly excited with a 514 nm argon ion laser.

On the contrary, the $E_A AC$ at times < 2 x 10^{-4} s appear to be dominated by Cy5 cis-trans isomerization, dynamics which are unrelated to FRET behavior as can be judged from the negative sign in the $E_A AC$ (see Figure 3.7). A fit to the experimental data for the controls done with –L3L4TAR with no NC and –L1L2L3L4TAR with and without NC reveals relaxation rate constants k_r ~ 1.8 x 10^4 s^{-1}. These fluctuations are not anti-correlated.

A similar relaxation rate constant k_r ~ 1.2 x 10^4 s^{-1} was measured for the relaxation of the Cy5 intensity autocorrelation both in the presence of an oxygen scavenger enzyme solution, and in a buffer solution in equilibrium with air (see Figure 3.7). Since oxygen has no effect on the relaxation rate constant for the Cy5 intensity autocorrelation, we rule out that this relaxation is due to Cy5 singlet-triplet interconversions. The only other plausible mechanism is cis-trans isomerization[84,104,105].

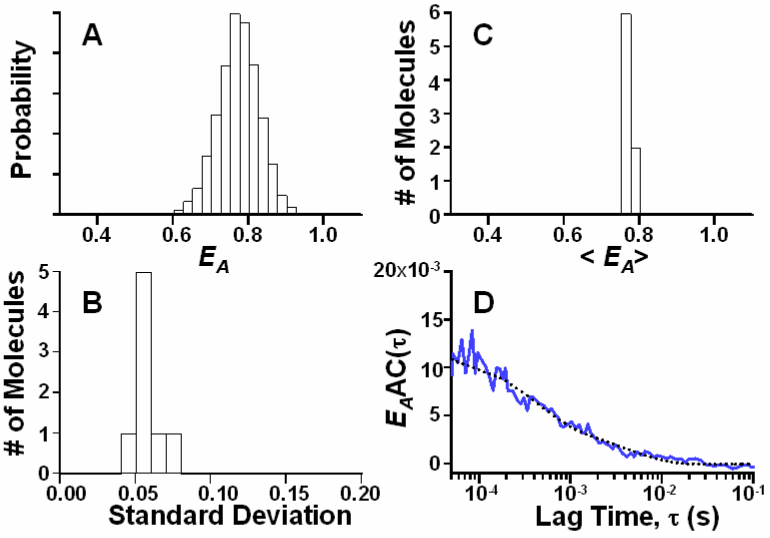


Figure 3.8: Data acquired with immobilized TAR DNA over a PEG coated glass coverslip.

(A) Ensemble distribution of $E_A(t)$ (constructed by superimposing the individual single-molecule distribution of $E_A(t)$ acquired with τ_B =10 ms). (B) Histogram of individual single-molecule $E_A(t)$ distribution standard deviations (σ_{E_A}). (C) Distribution of $<E_A>_{mol}$ values. (D) Ensemble E_AAC (in blue) and the fitting curve (black dot line) to the experimental data. All the data were acquired with 445 nM NC and 0.2 mM $MgCl_2$.

3.4. CONCLUSIONS

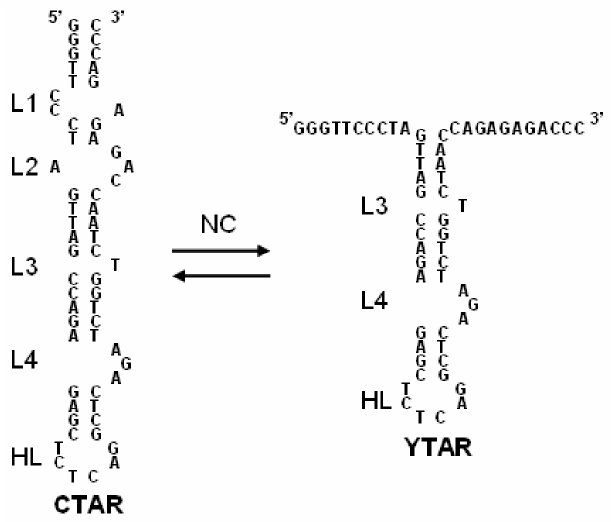
In order to elucidate the secondary structure of potential intermediates in the NC-catalyzed annealing process, we have undertaken the first time-resolved SM-FRET measurements on TAR DNA hairpins and hairpin mutants complexed with NC. The data were analyzed to determine the effect of NC complexation on the DNA end-to-end dynamics and end-to-end equilibrium distribution. The data show that NC shifts the equilibrium secondary structure of TAR DNA hairpins from a fully “closed” conformation to a “partially-open” conformation with the two terminal stems “open” or unwound and the other stems closed. Conformational equilibration (relaxation) occurs on a timescale (~ 1 ms) much shorter than the observation time (~ 4-20 sec) for each hairpin insuring equilibrium sampling in the observed single-molecule data. It can be argued that the observed partially-open TAR DNA/NC conformation is a critical intermediate in the NC catalyzed annealing mechanism of TAR DNA/RNA, since the open terminal stems possess 21 unpaired bases that are accessible for DNA/RNA annealing.

Chapter 4. Important Intermediates in the NC Protein Chaperoned TAR DNA/RNA Annealing

4.1 INTRODUCTION

4.1.1 Highly Heterogeneous Dynamics of C/Y Structures

In the previous chapter, SM-FRET was used to characterize the TAR DNA secondary structure in the presence of NC. It was determined that NC destabilizes the secondary structure in the 3'/5' terminal loop regions of the initially closed TAR DNA, termed the **CTAR**[67]. This process is illustrated in Scheme 4.1, in which the four bulge/loop regions are designated as L1-L4. The resultant system in the presence of NC is an equilibrium between CTAR and the partially open TAR DNA/NC complex, termed the **YTAR**.



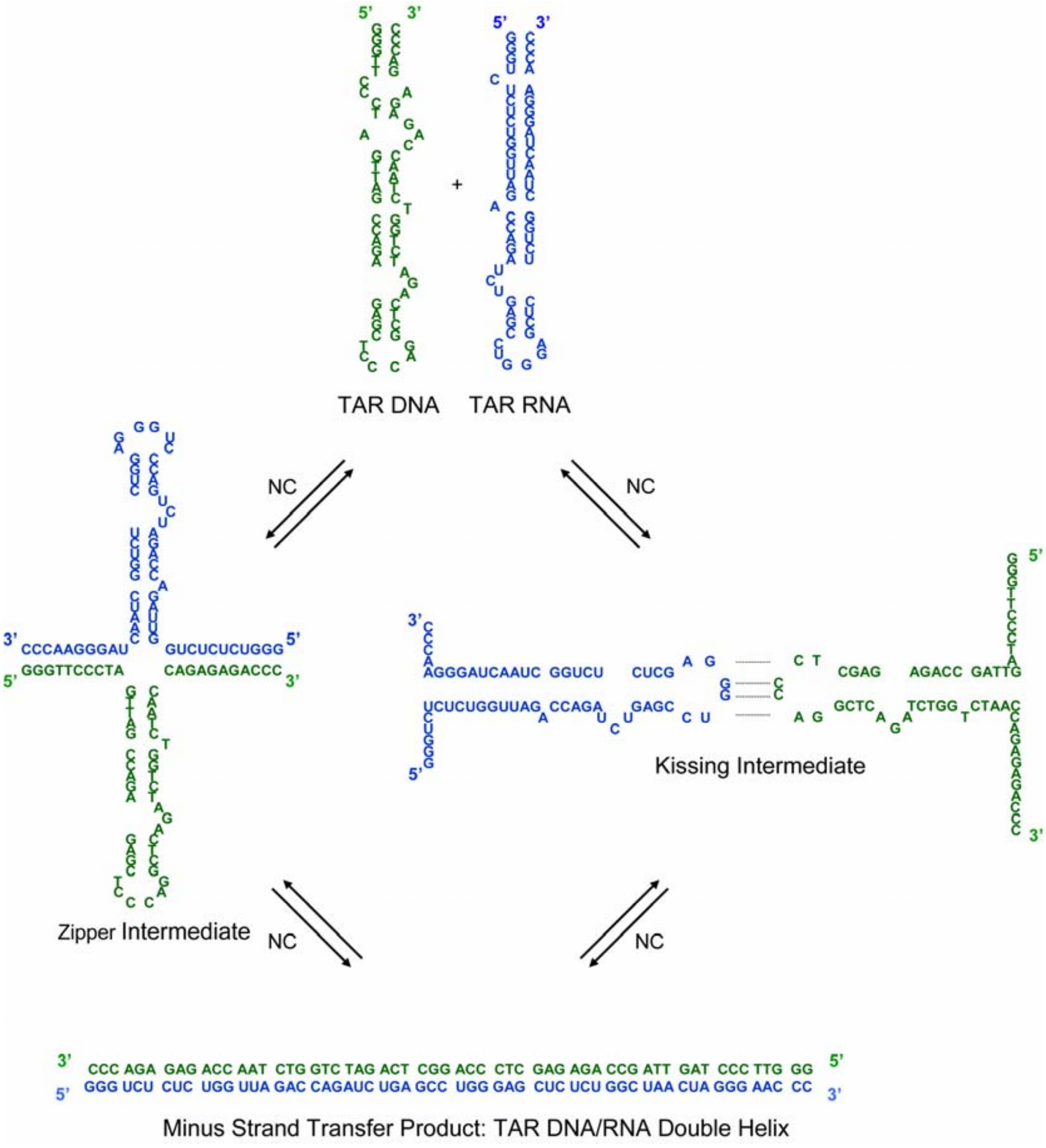
Scheme 4.1: NC induced the TAR DNA opening through the terminal L1L2 regions.
The partially open form in the presence of NC is termed the YTAR and the fully closed form is termed the CTAR [67].

Recently, an in depth SM-FRET study of the CTAR/YTAR equilibration dynamics was undertaken by Cosa et. al, revealing a strong dependence of the FRET dynamics on both the NC and Mg concentrations[106] leading to several new insights on the CTAR/YTAR interconversion mechanism. These single-molecule data indicate that NC induces a highly dynamic YTAR, as compared to the relatively conformationally static CTAR. The CTAR/YTAR dynamics were also observed to be significantly heterogeneous, indicating that the CTAR/YTAR interconversion is not a simple two-state equilibration but rather must involve multiple conformational intermediates and/or multiple reaction pathways[106]. The heterogeneity and the strong dependence of the dynamics on NC concentration is not unexpected since as many as 8 NC molecules can be simultaneously bound to TAR DNA[38].

4.1.2 Zipper Mechanism and Kissing Mechanism

We have suggested that the YTAR might serve as a nucleation center for minus-strand transfer[67]. In such a 'zipper' mechanism, nucleation occurs through either the 3' or 5' terminus in the NC-induced single-stranded form. In this model, the formation of such an intermediate increases the efficiency of forming a correctly annealed end product, and meanwhile decreases the probability of misfolded side products. The zipper mechanism is illustrated in Scheme 4.2 (left), where for brevity nucleation through both termini is shown.

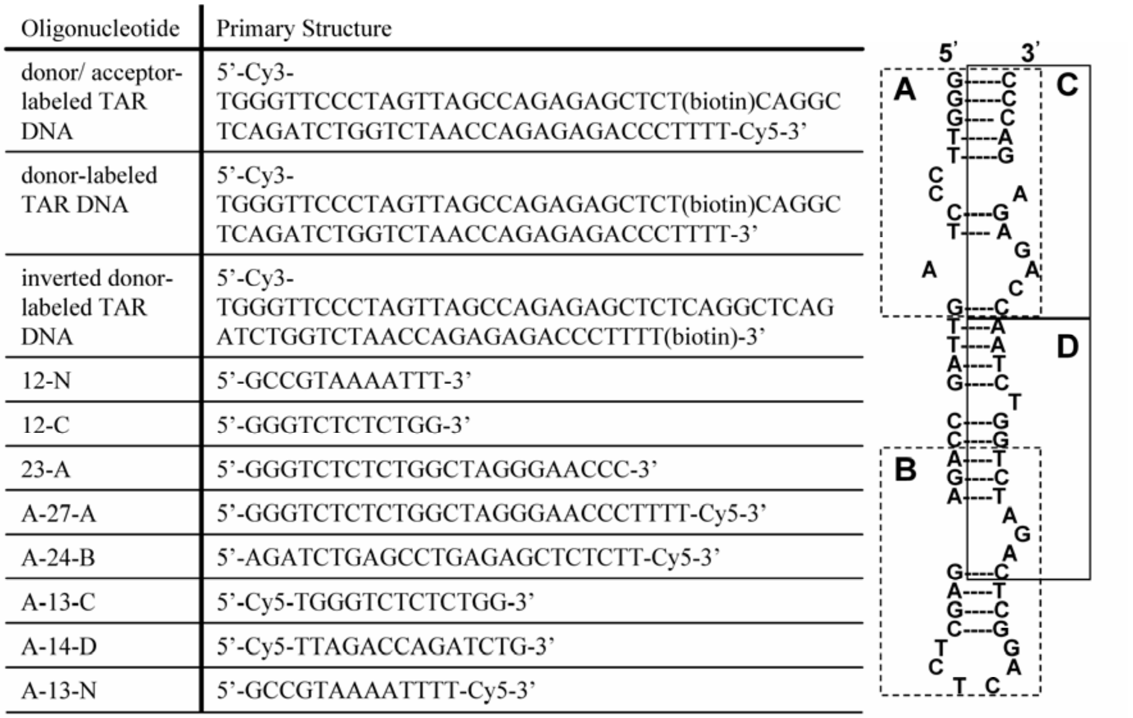
The second mechanism that has been proposed is the kissing mechanism[107], in which nucleation occurs through the complementary single-stranded loop sequences. NC-induced kissing complexes have been found to be important in stabilizing RNA dimerization in the viral envelope[108-112]. For minus-strand transfer this model involves the formation of inter-strand base pairs via loop-loop kissing, followed by formation of the extended DNA:RNA duplex. This mechanism also provides a specific nucleation site and decreases the likelihood of unfavorable side products. The formation of kissing complexes has been verified by NMR structural examinations of RNA/RNA interactions [113-115]. The kissing mechanism is illustrated on the right in Scheme 4.2.



Scheme 4.2: The two proposed mechanisms in the NC-chaperoned the annealing of TAR DNA and RNA.

4.2. DNA OLIGONUCLEOTIDES EMPLOYED IN THIS STUDY

The work in this chapter represents an effort to apply SM-FRET to characterize the two pathways that have been proposed for minus-strand transfer in HIV-1 reverse transcription. SM-FRET was used to monitor the NC chaperoned interaction between TAR DNA and various DNA oligonucleotides. The oligomers were chosen for their specific complementarity to regions of TAR DNA characteristic of each of the two pathways under consideration. The primary structure of each, including the respective location of the biotin, is listed in Scheme 4.3.



Oligonucleotide	Primary Structure
donor/ acceptor-labeled TAR DNA	5'-Cy3-TGGGTTCCCTAGTTAGCCAGAGAGCTCT(biotin)CAGGCTCAGATCTGGTCTAACCAGAGAGACCCTTTT-Cy5-3'
donor-labeled TAR DNA	5'-Cy3-TGGGTTCCCTAGTTAGCCAGAGAGCTCT(biotin)CAGGCTCAGATCTGGTCTAACCAGAGAGACCCTTTT-3'
inverted donor-labeled TAR DNA	5'-Cy3-TGGGTTCCCTAGTTAGCCAGAGAGCTCTCAGGCTCAGATCTGGTCTAACCAGAGAGACCCTTTT(biotin)-3'
12-N	5'-GCCGTAAAATTT-3'
12-C	5'-GGGTCTCTCTGG-3'
23-A	5'-GGGTCTCTCTGGCTAGGGAACCC-3'
A-27-A	5'-GGGTCTCTCTGGCTAGGGAACCCTTTT-Cy5-3'
A-24-B	5'-AGATCTGAGCCTGAGAGCTCTCTT-Cy5-3'
A-13-C	5'-Cy5-TGGGTCTCTCTGG-3'
A-14-D	5'-Cy5-TTAGACCAGATCTG-3'
A-13-N	5'-GCCGTAAAATTTT-Cy5-3'

Scheme 4.3: Primary structures of DNA oligonucleotides used in this study.

Right: A, B, C and D represent the regions of TAR DNA complementary to the DNA oligonucleotides listed on the left. Left: Oligonucleotides preceded by 'A' are Cy5-labeled. A-13-N and 12-N are oligonucleotides designed to be non-complementary to TAR DNA, as defined by fewer than 4 consecutive Watson-Crick base pairs.

Cy3/Cy5-labeled TAR DNA was used to monitor the effects of the NC and/or various DNA oligomers on the TAR DNA structure (see Scheme 2.3 A for the type I samples). Alternately, the acceptor dye was placed on the oligomers flowing into the flow cell for reacting with Cy3-labeled immobilized TAR DNA (see Scheme 2.3 A for the type II samples). This scheme yields information on the annealing between the two species.

In the various annealing experiments presented below, the NC concentration was held constantly at 445 nM, two orders of magnitude above the saturation concentration, assuming an approximate ratio of 1 NC bound to per 7-8 nucleotides[28]. All the samples were made with PEG surface (see section 2.1.1.2 for the PEG surface preparation) and all the solutions contain buffer A and an oxygen scavenger system (see section 2.13 for the details).

4.3 RESULTS AND DISCUSSION

4.3.1 Dynamics of TAR Structure under the Influence of Various Oligonucleotides

Columns 1 and 2 of Figure 4.1 summarize the results of a SM-FRET analysis of donor/acceptor labeled TAR DNA alone and in the presence of NC, respectively. The individual single-molecule E_A histograms (obtained from the single-molecule E_A trajectories) can be combined in an 'ensemble' histogram. In this case, the term 'ensemble' does not imply that the E_A values have been ensemble averaged. The resultant ensemble E_A histograms, gathered from 20-100 single molecules and comprised of thousands of occurrences, are normalized to a common scale. Ensemble E_A histograms for Cy3/Cy5-labeled TAR DNA in the absence of NC are shown in column 1 Figure 4.1. The data are presented at three different τ_B = 5, 25, and 250 ms. The near unity E_A value for Cy3/Cy5-labeled TAR DNA in the absence of NC indicates that it is in its CTAR form, shown in Scheme 4.1. The 5-ms τ_B data are broadened primarily by shot noise, but these effects are virtually eliminated once the data have been binned into 250-ms widths. The small broadening in the distribution is due to experimental errors. The ~1E_A distribution with narrow width and zero E_AAC amplitude support the model of CTAR as a static system with intact secondary structure.

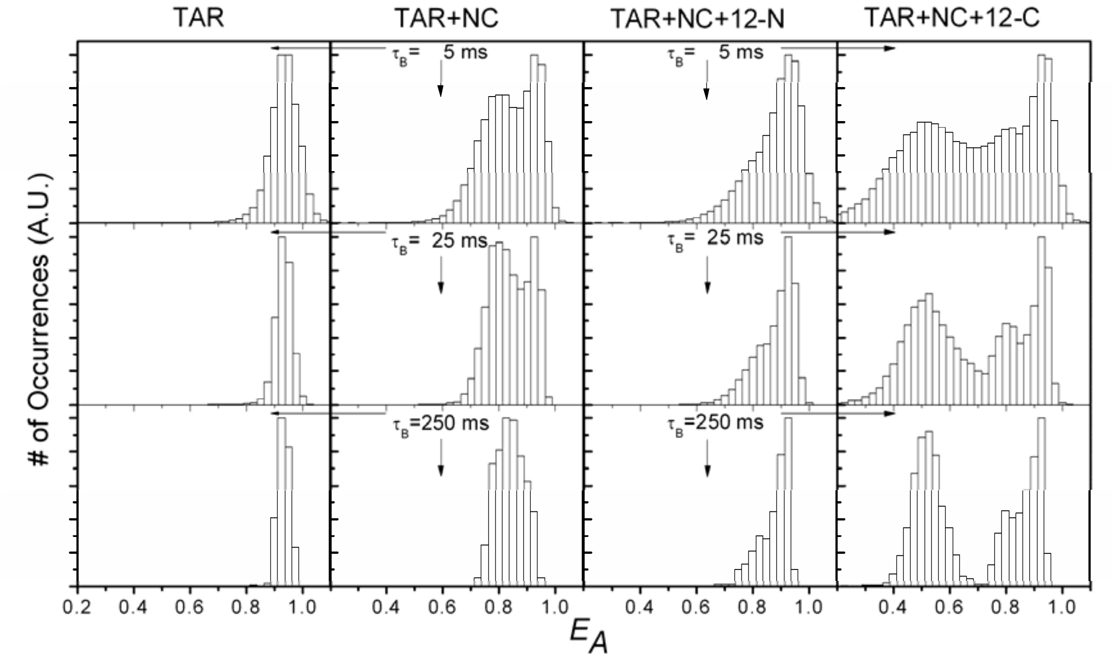


Figure 4.1: Ensemble E_A histogram of Cy3/Cy5-TAR DNA at three different τ_B under different reaction conditions.

Conditions for each column are: Column 1: with no addition of NC; Column 2: with the addition of 445 nM NC; Column 3: in the presence of 445 nM NC and 100 nM 12-N oligomer; Column 4: in the presence of 445 nM NC and 100 nM 12-C oligomer. All experiments were performed at 2 mM $MgCl_2$.

In contrast, when 445 nM NC is added (Figure 4.1, column 2) the histogram of E_A values becomes bimodal, indicating a dynamic distribution of both CTAR and YTAR. The results previously described in the Chapter 3 shows that NC-induced structural transformation between CTAR and YTAR[67] occurs on a millisecond time scale. At τ_B = 250 ms, the smeared-out bimodal E_A distribution (as compared to the E_A distribution at τ_B = 5 and 25 ms) in the presence of NC indicates that the dynamics of C/Y interconvesion is faster than 250 ms.

The critical dependence of NC activity on the TAR DNA secondary structure can be observed by introducing a single-stranded noncomplementary oligonucleotide to the system. The presence of the noncomplementary oligonucleotide, 12-N, at high concentration (100nM), only affects the relative distribution of CTAR and YTAR as indicated by the histograms shown at 3 different τ_B in column 3 of Figure 4.1. The ensemble E_A values show that the equilibrium of the TAR structure has shifted from the YTAR towards the CTAR. These results suggest that the effect of 12-N is to reduce the activity of NC. NC has a high binding efficiency for both TAR DNA and for the single-stranded 12-N. The lower effective concentration of NC results in increased occurrence of the CTAR[67,106].

The situation is different for the case of the complementary oligomer, 12-C. An additional feature ($E_A \sim 0.5$) is present in the E_A distribution shown in column 4 of Figure 4.1, which corresponds to the binding between 12-C and TAR DNA. Figure 4.2 shows that the process, however, is reversible on a ~10 s time scale (see further results on A-13-C, discussed below). Retention of separate species at τ_B = 250 ms indicates that the lifetimes of the bound and unbound systems of 12-C with TAR DNA are much longer than 250 ms. Indeed, the lifetime of the bound and unbound states obtained from fitting the on-off histograms with single exponential curve are 10 and 5 seconds respectively (see Figure 4.2).

Additionally, a similar shift from YTAR to CTAR can be observed as described for 12-N above. The cause is similar: competition for NC by the single-stranded 12-C decreases the activity of NC towards destabilizing L1L2. These results indicate that nucleation with a complementary oligonucleotide can occur at the 3’ end of the YTAR, providing support for the zipper pathway.

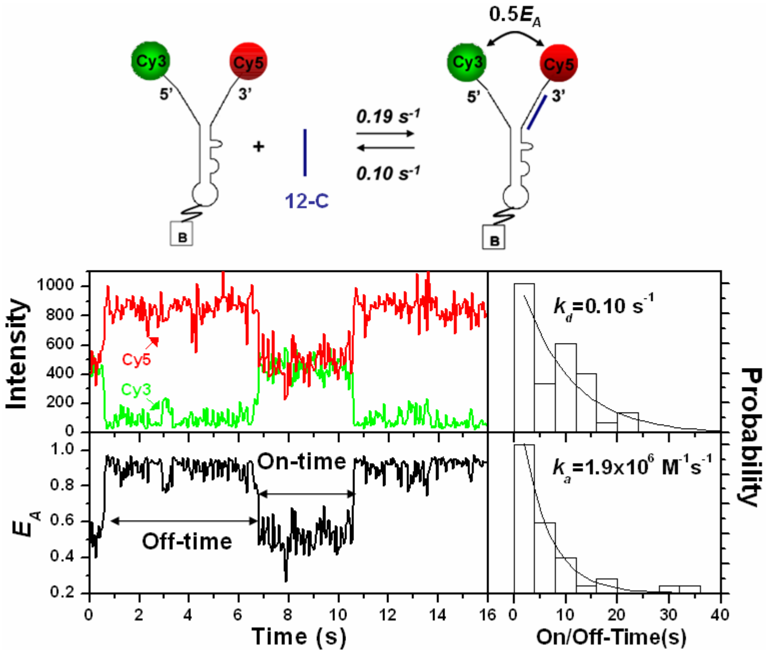


Figure 4.2: The on/off-time histograms revealing the dissociation/association rate constant of TAR DNA and 12-C short oligomers.

Left: An intensity (top) and E_A (bottom) time trace of a Cy3/Cy5-labeled TAR molecule annealing with 12-C short oligomers at 100 nM 12-C, 2 mM $MgCl_2$ and 445 nM NC. The time trace shows that TAR DNA is switching between an on (association) state and an off (dissociation) state. *Right*: An on-(top) and off-(bottom) histogram constructed by many on (E_A ~0.5) and off (E_A ~0.9) events from the same reaction are shown. The association rate of the TAR and 12-C (~10^6 M^{-1}s^{-1}) is similar to the annealing rate of TAR DNA/RNA obtained by the gel studies[116].

The dynamic equilibrium for the binding of shorter oligomers to TAR DNA is demonstrated in Figure 4.3. Here, the E_A histograms for the reversible binding of 12-C to Cy3/Cy5-labeled TAR DNA are compared for two different oligomer concentrations. In each case the histogram indicates the presence of three distinct species corresponding to CTAR, YTAR, and the bound 12-C/TAR DNA complex. The area of each peak in the histogram reflects the relative concentration of each form at equilibrium. When the 12-C concentration is increased to 100 nM, populations of bound complexes and CTAR are increased relative to the YTAR. This is due both to a higher 12-C concentration induced increase in yield, in the case of the bound form, and to the decreased activity of NC, in the case of the Y and C forms.

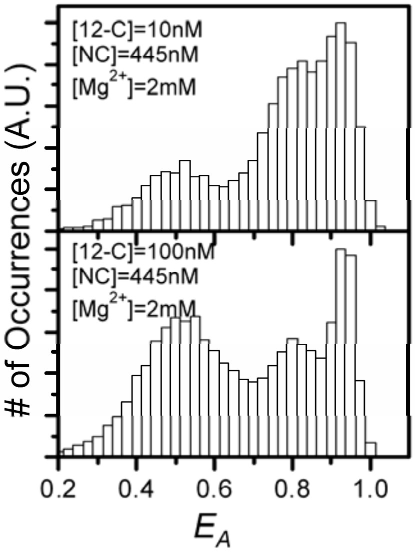


Figure 4.3: Ensemble E_A histograms (τ_B = 10 ms) of Cy3/Cy5-immobilized TAR DNA in the presence of 12-C and NC.

4.3.2 Annealing of TAR DNA with Various Oligonucleotides

To further examine this process, the interaction between Cy3-labeled TAR DNA and Cy5-labeled A-27-A (Scheme 4.3) was investigated. This construct was designed to mimic zipper nucleation at both termini. In contrast to the results obtained with 12-C, the reaction between TAR DNA and A-27-A occurs irreversibly within the time frame of the experiment, and results in a narrow distribution of E_A at ~1 (Figure 4.4 *A*). The high E_A is consistent with the structure illustrated (Figure 4.4 *A*, left) in which the two dyes are in close proximity due to complete annealing. An undetectably small E_AAC amplitude was observed indicating a static, fully annealed product (i.e. i.e. < 5x10^{-4}). These results provide support for a stable structure that is an excellent candidate for an intermediate occurring by the zipper mechanism.

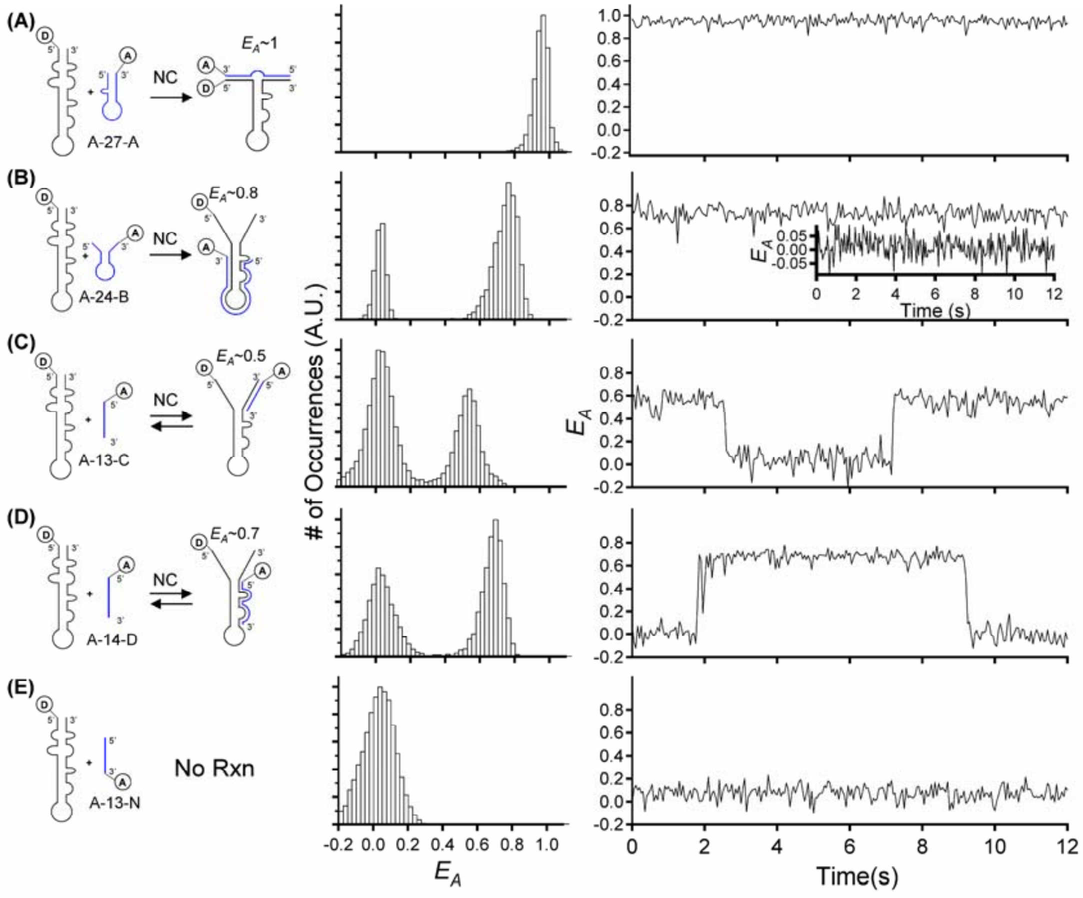


Figure 4.4: Annealing of various short oligonucleotides with the TAR DNA.

(A-E) Ensemble E_A histograms *(Middle)* and representative single-molecule E_A time trajectories *(Right)* are shown for annealing reactions of regular Cy3-TAR with Cy5-oligonucleotdies shown at left in the presence of 445 nM NC and 2 mM $MgCl_2$. The data are presented using τ_B = 50 ms. For reaction B, the inset on the right shows that no observable binding occurs in a minor population of the Cy3-TAR.

When the analogous experiment was performed using a DNA oligonucleotide designed to interact via the kissing mechanism, A-24-B, the formation of a stably annealed product was also observed (Figure 4.4 *B*). These data indicate that the kissing mechanism is also viable. However, the formation of the duplex via the kissing interaction required a longer time relative to the zipper construct (see below for further discussion), and in fact did not react to completion within the experimental observation time (~4 hours). This is indicated by the retention of unreacted Cy3-only emission observed in the E_A histogram in Figure 4.4 *B* (center column).

Representative single-molecule E_A trajectories, E_A(t), for reacted and unreacted molecules are shown in the right column and its inset. This reaction yield, however, reached > 90% when the inverted TAR was applied. The incomplete annealing, therefore, were resulted from the perturbation of the immobilization (see Figure 4.5). The time scale for the formation of the kissing-loop intermediate was similarly slow when inverted TAR DNA was used. The $<E_A>$ of ~0.8 for the annealed product is consistent with an increased distance between the two dyes, as illustrated (Figure 4.4 *B*, left). Like the zipper reaction mentioned above, the reaction was irreversible within the experimental time frame. However, both the relatively broad E_A distribution and a relatively large E_AAC amplitude (~10^{-3}) suggest that the resultant annealed structure is highly dynamic, compared to the product formed when TAR DNA was annealed with the zipper mimic discussed earlier. This is most likely due to residual fluctuations in the un-annealed 3' and 5' termini of TAR DNA, as observed earlier.

In contrast to the irreversible reactions observed in zipper and kissing reactions at 2 mM $MgCl_2$ and 445 nM NC, reversible binding was observed when SM-FRET was performed using shorter acceptor-labeled oligomers complementary to either the L1L2 region (A-13-C) or the L3L4 region (A-14-D) at the same condition. The ensemble E_A

histograms for each reaction are shown in Figure 3 *C* and *D*, respectively. Single-molecule trajectories in the right column for the nucleation of these shorter oligomers indicate binding reversibility, similar to the results described earlier for 12-C (Figure 4.1). Thus, in the ensemble E_A histogram for each reaction, the relative distribution of bound and unbound states reflects the equilibrium process. The K_{eq} for the reactions shown in Figure 3 *C* and *D* were calculated to be 0.62 and 1.19, respectively. The ensemble E_A histograms for these experiments (shown in the center column of Figure 4.4 *C* and *D*) yield evidence that in the presence of NC, the annealing can occur at any nucleation site along the TAR DNA hairpin. This is despite previous observations that the secondary structure of TAR DNA is not measurably weakened in the presence of NC in any region other than L1L2 [67]. These experiments also verify that the annealing processes are not perturbed by dye location. Non-specific binding can be ruled out because no reaction was observed when acceptor-labeled noncomplementary A-13-N was used, as shown in Figure 4.4 *E*.

Although as shown in Figure 4.4 A the annealing of zipper mimic, A-27-A, with TAR was irreversible, the reversibility of the zipper annealing was later observed to be dependent on the $MgCl_2$ and NC concentrations. Figure 4.5 shows the single-molecule E_A time trajectories of this annealing reaction performed at 445 nM NC and 2 different $MgCl_2$ concentrations. In these figures, each colored line represented an individual single molecule E_A time trajectories obtained from analyzing multiple confocal images acquired at various times. At 2 mM $MgCl_2$, the annealing of the zipper mimic with the TAR led to the formation of stable complexes. This was supported by the data in Figure 4.5 A showing that after the reaction time reached 2000 seconds, all the molecules went from $0E_A$ (free TAR) to $1E_A$ (TAR/A-27-A complexes) and stayed at $1E_A$ for the rest of the reaction. On the contrary, Figure 4.5 B shows that at 0.2 mM $MgCl_2$, the annealing of the A-27-A with the TAR led to a relatively unstable complex supported by the individual molecules constantly fluctuating between *0E_A* and *1E_A* through the whole reaction. Thus, the data demonstrated that the reversibility/reversibility of the annealing reaction is mediated by the $MgCl_2$ due to the effect of $MgCl_2$ on stabilizing the secondary structures of DNA duplexes[117].

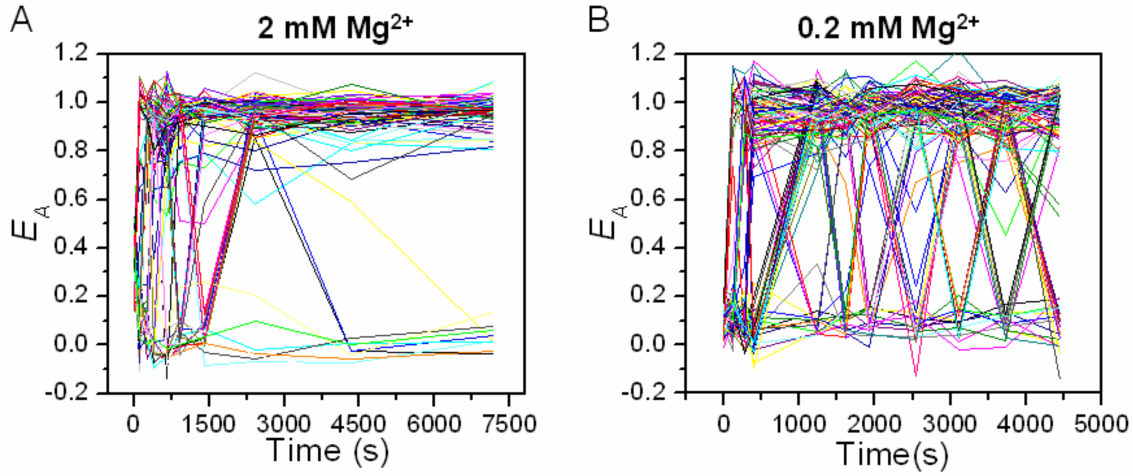


Figure 4.5: Single-molecule E_A time trajectories for the annealing of the zipper mimic, A-27-A with Cy3-TAR DNA.

The reactions were performed using regular Cy3-TAR in the presence of 445 nM NC, 10 nM A-27-A and (A) 2 mM and (B) $MgCl_2$.

The results presented in Figure 4.4 demonstrate that annealing can be initiated at multiple nucleation sites on the TAR DNA hairpin. Also, the relative yield of the kissing product versus the zipper product observed at room temperature suggests that the kinetic pathways for each type of intermediate may be different. In order to evaluate this hypothesis in more depth, the kinetics of formation of each type of product was measured, as shown in Figure 4.6. As can be clearly observed, the zipper product was formed more rapidly than the kissing product, with reaction lifetimes of 7 minutes, and 37 minutes, respectively. The analogous two reactions were also performed using inverted donor-labeled TAR DNA, yielding first order lifetimes of < 3 minutes and 21 minutes for the zipper and kissing products, respectively. (It needs to be noted that the ~ 3-min lifetime is close to the instrument response time which is limited by the confocal scan time per image, i.e. 2 min in these cases.).

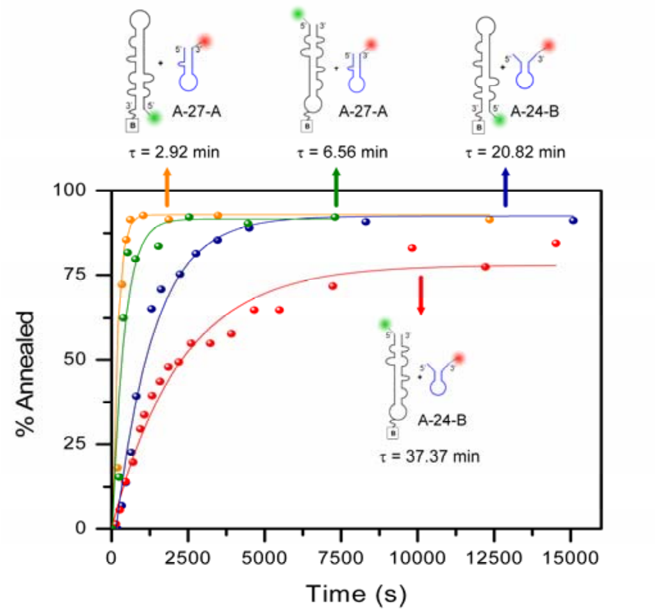


Figure 4.6: Kinetic curves for the annealing rates of the zipper mimic, A-27-A, vs. the kissing mimic, A-24-B, with Cy3-TAR DNA.

The reactions were performed using both regular and inverted Cy3-TAR in the presence of 445 nM NC and 2 mM $MgCl_2$. The oligonucleotide concentrations were 10 nM in all cases.

This kinetic results are consistent with our previous observation that in the presence of NC, the secondary structure in the L1L2 regions of TAR DNA is disrupted [67,106], allowing for more rapid annealing by the zipper mechanism. Interestingly, the kissing product formation lifetime was faster when the immobilization site was moved away from the hairpin loop region. This helps to confirm that the kissing product forms through inter-loop/loop interactions, because biotin immobilization in the hairpin loop would be expected to inhibit loop-loop interactions.

These results demonstrate that the zipper product is the more efficient pathway for the annealing of TAR DNA with DNA oligonucleotides. It is not the only pathway, however, and several types of evidence suggest that the kissing intermediate should also be considered as a candidate for TAR/TAR annealing[107,111]. Also, MFOLD predicts that the kissing intermediate is more thermodynamically stable than the zipper intermediate [118-121].

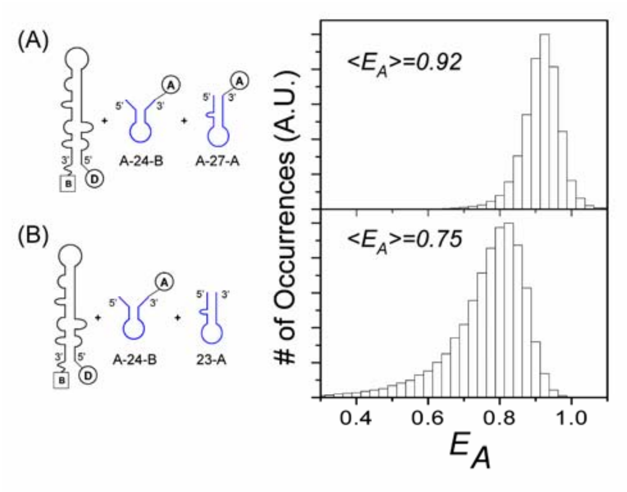


Figure 4.7: The competitive binding experiments using two oligonucleotdies complementary annealing with different regions of the TAR DNA.

Ensemble E_A histograms of annealing of Cy3-TAR with A-27-A and (A) A-24-B or (B) 24-B simultaneously at $[Mg^{2+}]$ = 2 mM and [NC] = 445 nM. The data are presented using τ_B = 10 ms.

The competition between kinetic and thermodynamic reaction control was tested using a competitive annealing experiment, in which both zipper and kissing intermediates were allowed to react simultaneously with TAR DNA. When both Cy5-labeled A-27-A and A-24-B were reacted with Cy3-labeled TAR DNA, the E_A histogram, shown in Figure 4.7 *A*, suggests that the zipper product dominates. The results of a control experiment (Figure 4.7 *B*), in which only the kissing oligomer was Cy5-labeled, suggests that both types of products were formed, however. Thus, although the zipper product forms more rapidly, as the kinetically favored pathway, the kissing product forms as well, consistent with its predicted thermodynamic stability.

4.4 CONCLUSIONS

The mechanism of NC chaperoned TAR DNA/TAR RNA annealing has been examined using TAR DNA and DNA oligonucleotides mimicking the potential annealing intermediates. In all cases, the previously reported Y structure of TAR DNA is preserved in the presence of NC and various oligonucleotides. It was found that annealing occurs through both zipper and kissing intermediates, although kinetic experiments indicate that the zipper product formation occurs more rapidly. In competitive reaction experiments, both types of intermediates are in fact observed. The results support a model of minus-strand transfer in which NC promotes the formation of multiple intermediates through which the fully formed product may occur.

Chapter 5. New Insights on the Role of Nucleic Acid/Protein Interactions in Chaperoned TAR DNA/RNA Annealing

5.1 INTRODUCTION

In the previous Chapter, SM-FRET measurements was employed to study the reversible annealing of various short DNA oligonucleotides to TAR demonstrating that the annealing can go through two potential nucleation sites, i.e. the “zipper” and “kissing” nucleation[106]. In the “zipper” mechanism[106] by analogy with the process of starting to zipper a coat, the annealing of TAR with short oligomers was initiated at the terminal single-stranded L1L2 region of the “YTAR” induced by NC. In the kissing mechanism, the annealing initiated at the single-stranded hairpin loop region of the “YTAR” was observed. The resultant products associated with both mechanisms were observed in the presence of NC, and the kinetics of formation of both types of products were also measured[106]. These studies lead to the proposal that the annealing of full TAR DNA and TAR RNA or complementary TAR DNA (cTAR) can go through the “zipper” and “kissing” mechanism (see Scheme 4.2 for the detailed information).

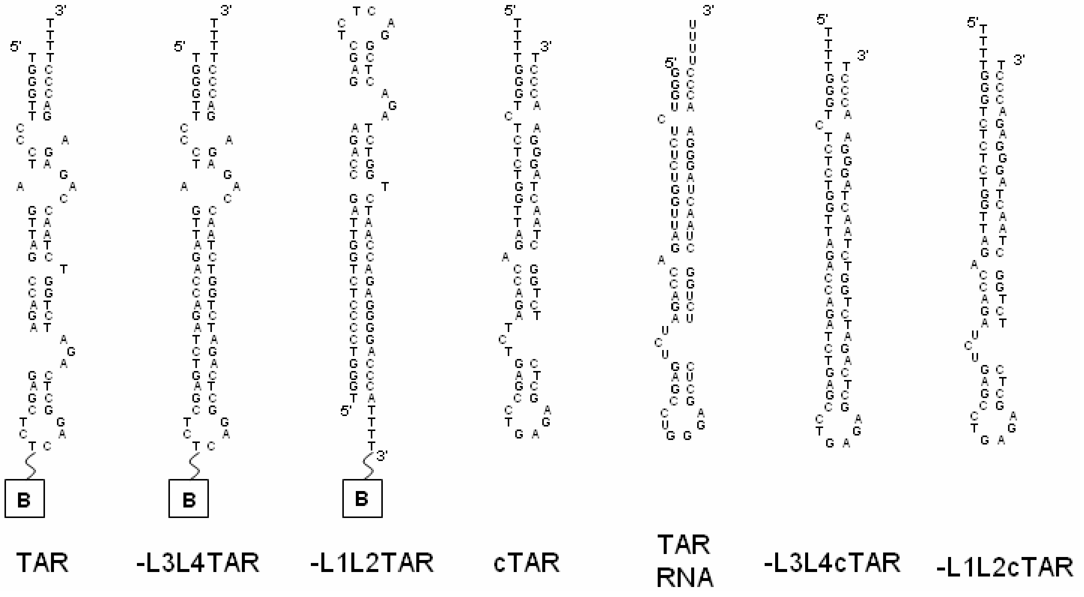
In this Chapter, SM-FRET kinetic measurements were performed to investigate the mechanism of the annealing kinetics of immobilized TAR DNA to full-length complementary non-immobilized TAR RNA or cTAR to examine the zipper and kissing pathways. We directly observe a diverse set of annealing intermediates including single cTAR hairpins coated by multiple NCs and large-scale NC/nucleic-acid aggregates containing many thousand RNA, DNA and NC molecules. This tremendous structural diversity is shown to make these reactions unsuitable for analysis by the usual “homogeneous” reaction mechanistic approaches that use for example the “order” of a reaction to help identify intermediates and transition states for a reaction. Fortunately, by employing a combination of SMS techniques, and by simultaneously controlling large scale aggregation, we have been able to address several outstanding mechanistic issues for minus-strand transfer including, the degree of complexation and secondary structure of the reactants *in situ*, the identity of the nucleation sites of annealing, and whether the loop-loop kissing interaction[86,113,122-124] is an important interaction in the mechanism.

5.2 OLIGONUCLEOTIDES USED IN THIS STUDY AND FCS MEASUREMENTS

5.2.1 DNA/RNA Oligonucleotides Employed in this Study

Annealing reactions were investigated between nearly all of the pair-wise combinations of immobilized and non-immobilized hairpins that are shown in Scheme 5.1, including the HIV-1 WT pair, TAR/TAR RNA. The various hairpins investigated include various mutants that were designed to probe how the kinetics depends on the presence or absence of internal bulges in the hairpins. Unless otherwise noted the annealing experiments employed NC concentrations = 890 nM and nucleotide/NC molar concentration ratios < 5 ensuring saturation binding of NC to the hairpins[16,19,29].

Immobilization of the TAR hairpins on a biologically compatibilized coverslip located in a flow chamber with biotin-streptavidin approach was employed (see Section 2.1.2) in order to keep one of the annealing reactants stationary and thereby allow for SM-FRET measurements at specific brief time intervals over the duration of the reaction, i.e. typically hundreds to thousands of seconds. The experimental design used a multiple syringe pump flow system allowing for rapid mixing of the NC with non-immobilized hairpin and then rapid delivery of this solution to the flow cell (see Section 2.1.3). This procedure effectively suppressed formation of large NC/nucleic acid aggregates (in most cases). Aggregation has previously reported to be a serious obstacle to making direct in-vitro measurements of the annealing kinetics for NC structures that contain the wild-type N-terminal 3_{10} helix domain[86,125]. In addition to WT TAR DNA/ TAR RNA annealing, various WT and mutant TAR DNA/DNA annealing reactions were also investigated as a mechanistic comparison and to take advantage of the easier access to DNA mutants and the smaller tendency of DNA to aggregate with NC[126-128].

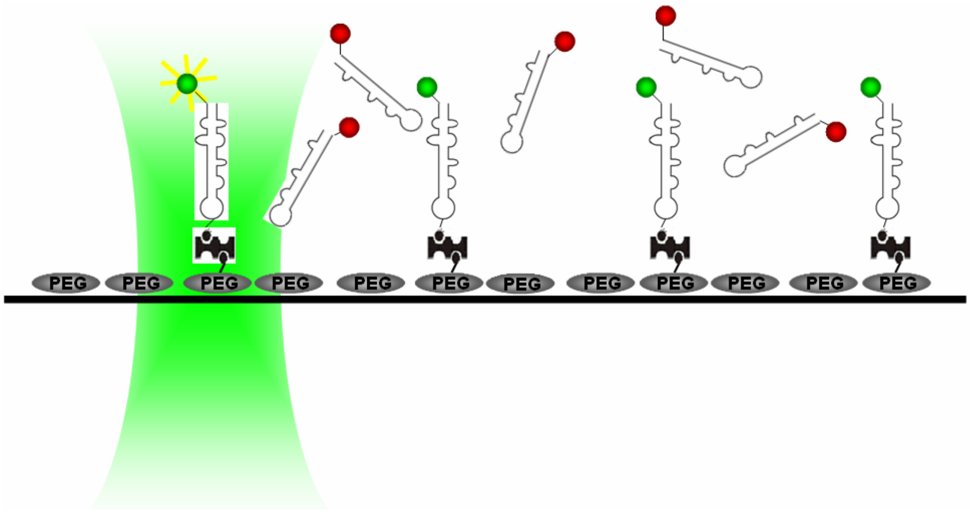


Scheme 5.1: Secondary structures of various oligonucleotides used in the annealing study.

The secondary structures are predicted by the mfold program (www.bioinfo.rpi.edu/applications/mfold/old/dna/). The first 3 hairpins with a biotin linkage represent the immobilized DNA hairpins (indicated as B) and the immobilized DNA. The rest 4 DNA/RNA hairpins represent the hairpins flow into the flow cells.

The immobilized hairpins were labeled at the 5' end with a Cy3 dye (FRET donor dye) and the complementary non-immobilized hairpins were labeled at the 3' end with Cy5 (or Dylight dye in the case of TAR RNA) dye as an acceptor, with the type II sample configuration (see Section 2.1.2.2 and Scheme 2.3)[106]. Cy3/Cy5 is a well-established donor/acceptor dye pair allowing for an instantaneous measurement of the inter-hairpin

Cy3-Cy5 distance by SM-FRET[78,79,129]. Irreversible annealing kinetics were initiated by exposing a dilute immobilized sample of the Cy3-TAR to a "fresh" solution of Cy5-hairpins (4 non-immobilized hairpins as shown in Scheme 5.1) and NC at zero time, t = 0, in analogy to a stopped-flow experiment as shown in Scheme 5.2 for the NC catalyzed annealing of TAR to cTAR DNA. The fluorescence spots in the confocal images for the donor and acceptor channels were used to determine the acceptor and donor intensities, $I_A(t)$ and $I_D(t)$, to calculate the $E_A(t)$ (see Section 2.3).



Scheme 5.2: The type II sample combined with the flow system.

By measuring E_A for each hairpin at various times, t, after introducing the Cy5-hairpin solution, single-molecule E_A time trajectories (see Section 2.5, Scheme 2.8) were recorded to monitor the instantaneous distance between the 5' end of the immobilized Cy3-DNA and the 3' end of the Cy5-labeled complementary DNA/RNA. The time spacing between FRET points (i.e. confocal images) was varied during the experiment to minimize photobleaching. Typically, the time spacing was > 120 s, i.e. the time to record one confocal image. In certain cases, 10-ms resolved FRET trajectories were recorded for individual hairpins at specific times along the course of the reaction by positioning the stage on a specific immobilized DNA.

5.2.2 FCS Measurements and Data Analysis

A 633-nm cw laser was coupled and collimated into the microscope and focused via the microscope objective as previously described to carry out the FCS measurement. (in this case, the laser is not modulated by the AOM). The microscope objective focused the excitation light inside the solution (ca. 20 μm above the coverslip surface). The fluorescence emissions from Cy5-cTAR were then collected through the same objective, separated by a 50/50 beamsplitter and focused onto 2 APDs serving as pinhole to reject the out-of-focus fluorescence. The autocorrelation of the fluorescence signals was obtained by cross-correlating 2 signals split from the collected light to avoid the effect of the "afterplusing"[83]. For the detailed information of the layout of the apparatus, please see Section 2.2.

The correlation data was acquired by a real-time digital correlator ALV 5000/E correlation board. The FCS data of Cy5-cTAR in the absence or presence of NC (shown in Figure 5.4) are well-fit by a model based on a freely diffusing fluorophore in a 3D Gaussian focal volume. The correlation function can be express as follows (the same as Equation 1.6, all the variables τ_D, ω, p and τ_T has been described in Section 1.4.3):

$$G(\tau) = 1 + \frac{1}{N}(1 + \frac{\tau}{\tau_D})^{-1}(1 + \frac{\tau}{\omega^2 \tau_D})^{-1/2}[1 + \frac{p}{1-p}\exp(-\frac{\tau}{\tau_T})] \qquad (5.1)$$

The average number of cTAR molecules present in per cTAR/NC aggregate is determined by the value of G(τ = 0) which is inversely proportional to N.

5.3 RESULTS AND DISCUSSION

5.3.1 TAR/Complementary TAR Annealing

In the presence of both NC and WT Cy5-cTAR DNA in solution, the observed E_A trajectories exhibit discrete jumps from an E_A value near zero to a value near unity (see below Figure 5.1 A) at various times during the annealing reaction. With either or both of NC and complementary hairpins absent the E_A value does not vary from zero over the entire time scale of the experiments. Figure 5.1 A demonstrates that the discrete, "telegraphic" nature of the annealing trajectories is especially apparent in "kinetic" histograms of the SM-FRET E_A determined for specific time windows after the initiation of the reaction for an ensemble of reacting Cy5-species (see Figure 5.2). Thus, the SM-FRET kinetic results directly reveal, in a way that would not be possible with ensemble measurements, that the annealing reaction evolves to kinetic "stable-states", i.e. all other nucleic acid rearrangements must occur on a more rapid time-scale. The mean apparent FRET, $\langle E_A \rangle$, as a function of reaction time for an ensemble of immobilized hairpins is also shown (square data points in Fig. 5.5A). This single-molecule method for determining $\langle E_A \rangle$ is superior to ordinary bulk solution methods that are subject to artifacts such as mislabeled hairpins, which can easily and automatically be removed from the SMS data during analysis.

The number of reactant and product molecules, N_R and N_P, respectively were counted directly by using an E_A threshold of 0.5 to distinguish between reactants and products (as shown in Figure 5.1 A, the $0E_A$ and $1E_A$ corresponded to the reactant and product). These N_R and N_P data for TAR/cTAR annealing under saturated NC binding (green and blue solid lines in Figure 5.1 B respectively) were well-fit with a single exponential function (dashed-lines), which is the predicted kinetic behavior for an irreversible bimolecular reaction under the present pseudo-first order conditions for these non-immobilized reacting species. The best-fit parameters include a pseudo first order rate constant, k_ψ, and a percent annealed, the latter of which is close to the expected 100%. The single exponential behavior suggests the existence of a kinetic bottleneck for annealing. In chemical kinetics, a single bottleneck often implies a single well-defined transition state. In the case of the annealing reaction, the conclusion of a well-defined bottleneck applies to the aggregation free (homogenous solution) mode of the annealing reaction. A more complex mechanism that involves more than one kinetic bottleneck applies when aggregation is not controlled, see Section 5.3.2.

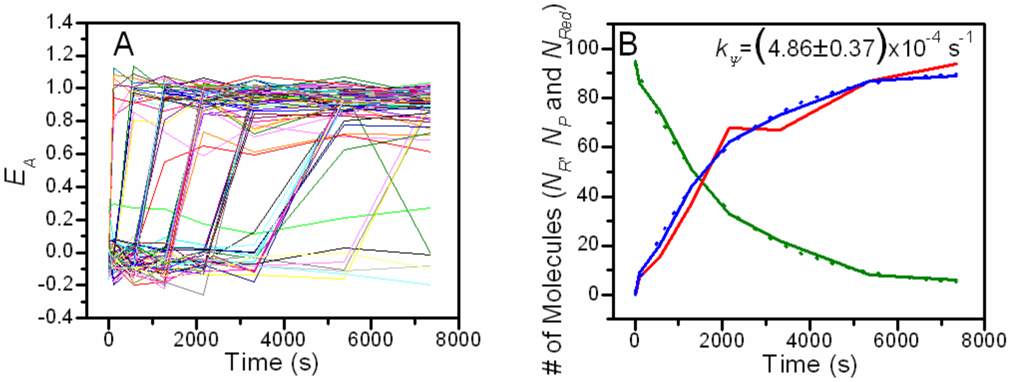


Figure 5.1: SM-FRET measurements of the TAR/cTAR annealing reaction.

(A) E_A time trajectories of single molecules for the TAR/cTAR annealing reaction at [cTAR] = 2nM, [$MgCl_2$] = 0.2mM and [NC] = 889 nM. Each colored line represents individual single molecules. (B) The number of reactant (blue, solid line), product (green, solid line) and Cy5 molecules (red, solid line) as a function of time and fit to the single exponential curve (corresponding dash lines) yielding a pseudo first order rate constant, $k\psi = (4.86 \pm 0.37)\times10^{-4}\ s^{-1}$ and 94% annealed product.

A distinct advantage of the scanning confocal method for SM-FRET is that it allows for synchronous detection of the Cy3 and Cy5 fluorescence intensity while rapidly switching the laser excitation wavelength between 514 nm and 633 nm, which are respectively the wavelengths that excite Cy3 and Cy5 (see Section 2.2.2 for the two-color alternating excitation setup). The 633-nm excitation induces only Cy5 fluorescence, which is a quantitative measure of the number of bound Cy5-labeled species, termed the N_{Red} at various times during the reaction, after calibration. The SM-FRET kinetic data in Figure 5.1 B shows that the number of associated cTAR species of any type (red curve) is close to the number of annealed duplexes (blue curve) throughout the annealing reaction, within experimental error. The slight deviations of N_P from N_{Red} are expected since the photobleaching of Cy5 dyes on individual molecules occurs during the reaction. Two procedures were used to make FRET-free determination of the number of Cy5-labeled hairpins, N_{cTAR} (or N_{Red}) associated with an immobilized hairpin during the annealing reaction (for the procedures to count the Cy5-labeled cTAR, please see Section 2.5). These data were compared with the number of annealed hairpins, N_P, determined by FRET, leading to the conclusion that N_{cTAR} (or N_{Red}) is equal to N_P within experimental error throughout the annealing reaction as shown by the red line in Figure 5.1 B.

This type of kinetic curve strongly demonstrates that a Cy5-labeled cTAR hairpin is not associated (i.e., co-localized) with each Cy3-immobilized TAR hairpin until the FRET change occurs, i.e., indicating an annealing event. In addition, it was possible to show that the on average a TAR hairpin is associated with a Cy5-cTAR hairpin less than 3% of the time before the annealing event. This value can be used as an upper limit for the concentration of the NC bound TAR/cTAR encounter complex, allowing for a calculation of a mean upper-limit of >70 nM for the dissociation constant of the NC bound encounter complex. The same measurement was also made with a non-complementary Cy5-labeled hairpin at 2 nM (for which an even more sensitive measurement could be made since annealing was blocked). For this latter case the dissociation constant of the putative NC bound complex was estimated to be >200 nM. Thus, these results show that the various non-immobilized hairpins in Figure 5.1 B are not substantially associated by NC to the various immobilized hairpins until actual annealing occurs. The same type of experiments with the other immobilized hairpins mutants shown in Scheme 1, gave analogous results. An exception to this behavior occurs during large scale NC induced nucleic acid aggregation, see Section 5.3.2.

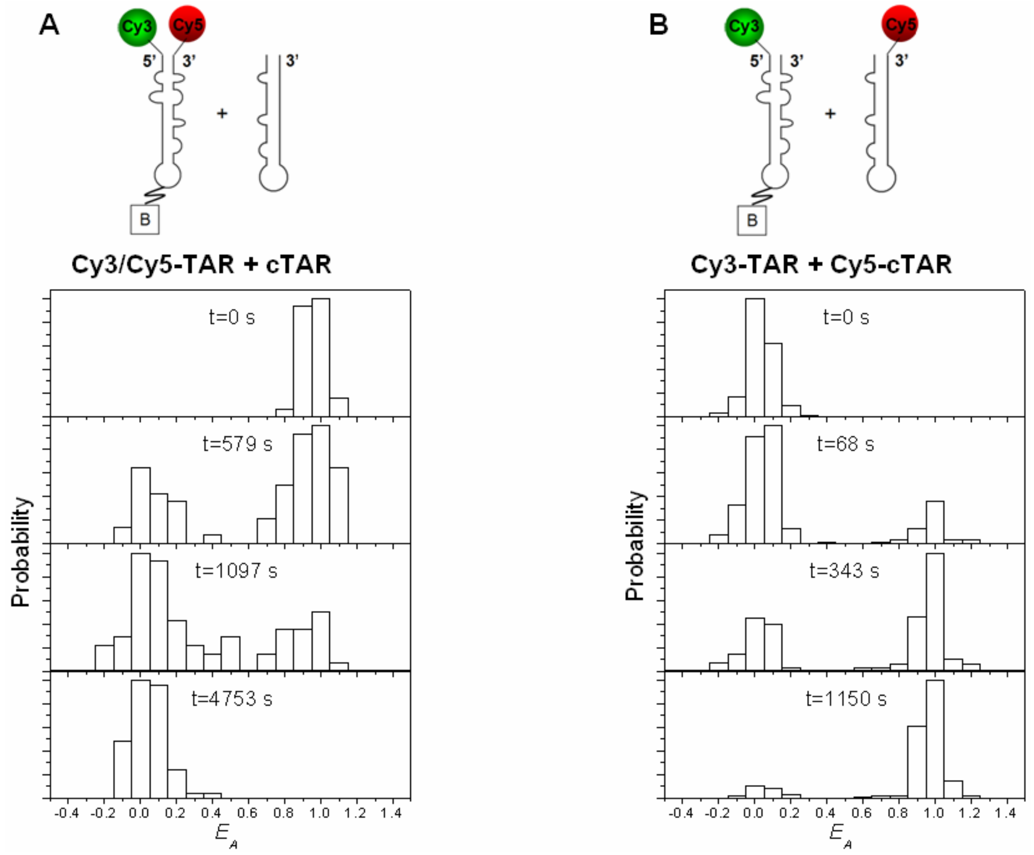


Figure 5.2: Ensemble $E_A(t)$ histograms of TAR DNA annealing with cTAR as a function of time.

$E_A(t)$ histograms constructed at different time after initiation of the NC-induced annealing of TAR with cTAR The reactions were performed using (A) Cy3/Cy5-TAR with cTAR and (B) Cy3-TAR annealing with Cy5-cTAR. The reaction condition is 2mM $MgCl_2$ and 889 nM NC and 10 nM cTAR.

Further information on the secondary structure and degree of aggregation of the reactant TAR hairpins was obtained by recording the E_A histogram of immobilized doubly-labeled 5'/3' Cy3/Cy5-TAR in the presence of both NC and the unlabeled complementary cTAR reactants. During the reaction the E_A histogram exhibited a major peak at 0.9 due to that the TAR is in equilibrium between the "C" and the "Y" structure at 2mM Mg (see Figure 5.2 A)[67,106], and a second peak at $E_A \sim 0$ assigned to the duplex product of annealing, which has a significantly increased spacing between the 3' and 5' ends. The presence of non-labeled cTAR under typical reaction (10 nM cTAR) conditions has no measurable effect on the $0.9E_A$ peak assigned to "YTAR" at the same condition with no cTAR (2mM Mg^{2+} and 889 nM NC), further demonstrating that the reactant TAR DNA hairpins are not significantly associated with cTAR under these conditions. Thus, a single "YTAR" coated by multiple NC is the dominant form for the reactant in the NC induced annealing mechanism of TAR DNA.

5.3.2 NC/Nucleic Acid Aggregation

Under conditions in which large scale NC/nucleic acid aggregates are present, i.e. at high cTAR (or TAR RNA) concentration, the observed rate constants for NC induced annealing for the various hairpins fluctuated greatly from trial to trial. This is reflected in Figure 5.3 and Table 5.1 by the extraordinary large standard deviations for high cTAR concentration for TAR/cTAR annealing. The annealing kinetics fluctuates from run to run presumably due to the aggregation. The bimolecular rate and the relative standard deviation of the rates obtained from the total 10 trials are $(2.72 \pm 1.2) \times 10^5$ ($M^{-1}s^{-1}$) (as shown in the Table 5.1 as well) and 44%.

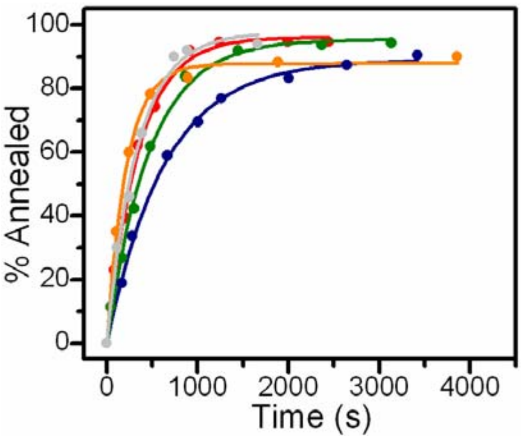


Figure 5.3 5 trails of kinetic curves of TAR:cTAR annealing chaperoned by NC at 2mM $MgCl_2$ and 880nM NC.

Each colored lined represents one annealing trial performed on different day. The total number of trial for this particular annealing reaction performed at this condition is 10.

A number of hypothetical origins for the rate fluctuations were ruled out caused by the following reasons: (1) the concentration fluctuation of NC or cTAR caused by NC or cTAR stuck to the tubing of the flow system and NC or the nucleic acids adsorbed onto the PEG surface or (2) temperature variations due to the temperature fluctuations nearby the sample or the motor of the syring pump heating up the syringe over time. Ultimately, the reaction rate fluctuations were assigned the concentration fluctuation of NC or cTAR due to the aggregation of the cTAR with NC in large molecular aggregates that contain hundreds to thousands of cTAR and NC molecules. The number of cTAR molecules per aggregate was estimated from the fluorescence burst intensity for individual aggregates (see Figure 5.4), using the known fluorescence intensity for individual hairpins at the same excitation intensity as a calibration.

Direct evidence for NC induced aggregation was obtained by recording time resolved fluorescence intensity curves, *I(t)*, for Cy5-labeled cTAR in solution. Here the scanning stage was positioned over a region which was free of immobilized hairpins, and the fluorescence induced by 633-nm excitation was recorded for a relatively high concentration of Cy5-cTAR in solution in the excited volume of the confocal, in the absence (blue) and presence (red) of NC, see Figure 5.4. In the absence of NC, I(t) only exhibits short time fluctuation due to photon shot-noise and concentration fluctuations that are consistent with the diffusion of single, non-aggregated Cy5-cTAR hairpins. In contrast, in the presence of NC, the I(t) curves show clear evidence of large scale nucleic acid aggregation. For example, intense blips are observed due to single aggregates

Table 5.1: The apparent second-order TAR:cTAR annealing rate constants at various NC, $MgCl_2$ and cTAR concentrations

$[Mg^{2+}]$ (mM)	[NC] (nM)	[cTAR] (nM)	k_a ($M^{-1}s^{-1}$)	[NC]/[nt]
2	889	5	$(3.64 \pm 0.34) \times 10^5$	2.8
		10	$(2.72 \pm 1.2) \times 10^5$	1.4
		20	$(3.21 \pm 0.31) \times 10^5$	0.7
1	889	10	$(6.66 \pm 0.47) \times 10^5$	1.4
0.2			$(8.37 \pm 0.048) \times 10^5$	
0.2	500	50	$(1.69 \pm 0.09) \times 10^5$	0.16
		83.3	$(9.16 \pm 10.9) \times 10^2$	0.09
		> 500	No Annealing	< 0.016
0.2	300	50	$(3.73 \pm 2.77) \times 10^4$	0.09
	< 200		No Annealing	< 0.063

The kinetics were all measured under pseudo first order conditions, where the [cTAR] >> [TAR]. The apparent second-order rate constant, k_a was estimated by dividing the pseudo first order rate constant by [cTAR]. In fact, the reaction is only second-order at high values of [NC]/[nt], see text for further detail. The solutions contained buffer A and an oxygen scavenger system and the reaction were performed at room temperature.

containing thousands of Cy5-cTAR. Also, a significant decrease in the steady-state concentration of the cTAR monomer is observed resulting from incorporation of cTAR in the NC assembled aggregates. The *I(t)* curves indicate that both hairpin monomers and aggregated hairpins are present simultaneously. Aggregation was observed to be especially severe for solutions with high hairpin concentrations, and more severe for RNA than DNA. These various observations suggest that the aggregation process is analogous to a nucleation controlled precipitation. By using high flow rates (freshly mixed solutions) it was possible to suppress formation of aggregates in the presence of NC for small nucleic acid concentrations. This was confirmed by fluorescence correlation spectroscopy (FCS) on Cy5-cTAR in the absence and presence of NC (Figure 5.4, inset). The FCS data in both cases are well-fit by the standard model (see Equation 5.1) with a best-fit diffusion constant.

The aggregation phenomenon was observed to produce huge fluctuations of the annealing rate at high cTAR and TAR RNA concentrations, even inducing a "stalling effect", i.e. no detectable reaction despite high cTAR and NC concentration (at 500 nM NC and 500 nM cTAR, see Table 5.1). Over this range the kinetic order of the reaction for cTAR, was observed to vary from first-order at ≤ 20 nM cTAR concentrations, to zero-order at concentrations in the 25-100 nM range, and then ultimately to actually exhibit negative orders when the reaction stalled at high cTAR concentration (i.e. 500 nM).

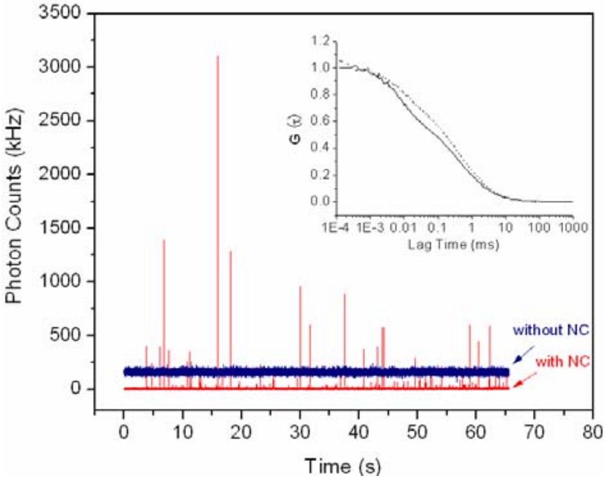


Figure 5.4: Representative *I(t)* curves detecting the emission of Cy5-cTAR DNA obtained by FCS measurement.

Blue lines are the intensity trajectory of Cy5-cTAR DNA (50 nM) without NC present, exhibiting a continuous signal of ~160 kHz (photon counts per second). Red lines correspond to Cy5-cTAR DNA (50 nM) with NC (500 nM), revealing intense blips due to aggregation and a decrease in the steady-state signal from monomer Cy5-cTAR DNA to ~4 kHz. (Inset) FCS curves of cTAR (5 nM) (solid line) and cTAR (5 nM) with NC (500 nM) (dashed line) showing increase of τ_D due to NC binding to cTAR. Here, τ_D is the diffusion time for cTAR or cTAR/NC complexes passing through the focal volume of the laser.

5.3.3 Annealing Rate Trends in the Absence of Aggregation

Aggregation, which appears to be kinetically controlled process, can be largely avoided for low cTAR and TAR RNA concentrations (i.e. < 10 nM). Under these conditions, analysis of WT and mutant TAR DNA reveals several mechanistically informative trends. The WT TAR/cTAR and TAR/TAR RNA (with L1-L4 preserved for both TAR and cTAR/TAR RNA) reactions are observed to be overall second-order, i.e. the pseudo first rate vs. cTAR is linearly proportional to [cTAR] (see Table 5.1 above).

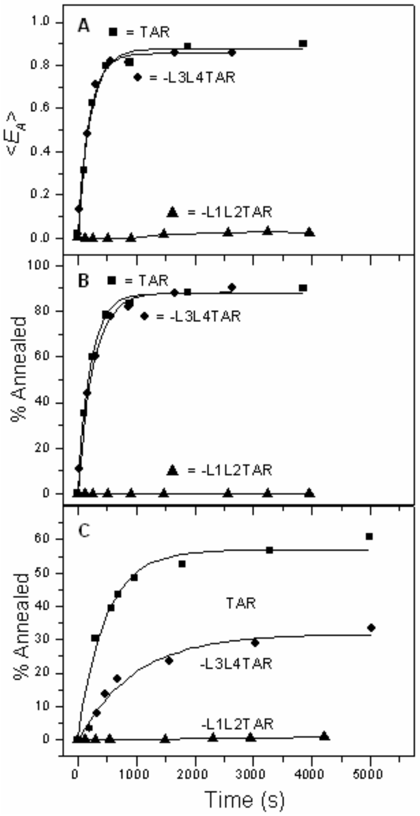


Figure 5.5: SM-FRET measurements of the annealing kinetics using various WT and mutant TAR DNA/cTAR (RNA) pairs.

E_A (A) and % annealed (B) vs. time for the annealing of cTAR with WT and mutant TAR DNA hairpins at 10 nM cTAR, 2 mM $MgCl_2$ and 890 nM NC. (C) % annealed vs. time for annealing of TAR RNA with various immobilized WT and mutant TAR DNA hairpins at 0.2 mM $MgCl_2$ and 890 nM NC. For TAR DNA, 10 nM TAR RNA was used, whereas for -L3L4TAR and -L1L2TAR, 25 nM TAR RNA was used.

At NC concentrations below 20 nM, the rate falls off precipitously with [NC] as expected for NC induced cooperative melting of TAR and cTAR, which is a key step in the annealing process. For -L1L2TAR/cTAR annealing the data clearly show that the L1L2 loops of TAR are required and in fact are sufficient for rapid annealing (see Figure 5.5A and 5.5B and Table 5.2 for the annealing rate constants). In contrast, removing the L3L4 loops has little kinetic consequence. Similar trends are observed for TAR/TAR RNA annealing. While removing the L3L4 loops does not stop the rapid initiation of annealing, it apparently shifts the annealing equilibrium toward reactants. The lower apparent annealing percentage for TAR RNA has been extensively investigated and has been assigned to several factors, including a incomplete of dye labeling (probably ~ 70%), errors in the synthesis producing a small concentration of TAR RNA hairpins with missing bases, and complications due to aggregation.

Table 5.2: The effect of loop removal on the rate constants of TAR DNA/cTAR annealing in the presence of 890 nM NC and 2mM $MgCl_2$

Non-immobilized DNA	Immobilized DNA	k_a ($M^{-1}s^{-1}$) Avg. ± sd. Single exponential fitting[c]
cTAR	TAR[a]	$(2.72 \pm 1.2) \times 10^5$
	-L3L4TAR	$(3.03 \pm 0.61) \times 10^5$
	-L1L2TAR	No Annealing
-L3L4cTAR	TAR	$(2.16 \pm 0.96) \times 10^5$
	-L3L4TAR	$(1.98 \pm 0.71) \times 10^5$

All the reactions were performed in buffer A and oxygen scavenger system at room temperature. [a]For TAR:cTAR annealing at different concentrations of cTAR and $MgCl_2$, please see Table 5.1 for the details. [c]For -L3L4TAR:cTAR annealing, a bi-exponential fitting curve can be obtained, which gives a fast rate, $(3.52 \pm 0.75) \times 10^5$ ($M^{-1}s^{-1}$) and a slow rate, $(5.33 \pm 2.71) \times 10^4$ ($M^{-1}s^{-1}$).

5.3.4 The Mechanism of TAR DNA/RNA Annealing promoted by NC

Based on the new information obtained herein a refined and more specific mechanism for the NC chaperoned annealing of TAR to cTAR or TAR RNA has been constructed, as shown in Scheme 5.3. This mechanism applies exclusively to the annealing reaction in the absence of large scale nucleic-acid/protein aggregates. The SM-FRET results demonstrate that the TAR reactant is predominantly a single NC coated, hairpin with a dynamic secondary structure, involving a partially-open "Y" shaped conformations, as shown in equation (i) of Scheme 5.3. Likewise, Figure 5.6 demonstrates NC also induced a partially melted "Y"-like structure of TAR RNA (and presumably as well as cTAR). The FCS data demonstrate that the complementary cTAR or TAR RNA reactants are single non-aggregated hairpins, i.e. except for the bound NC.

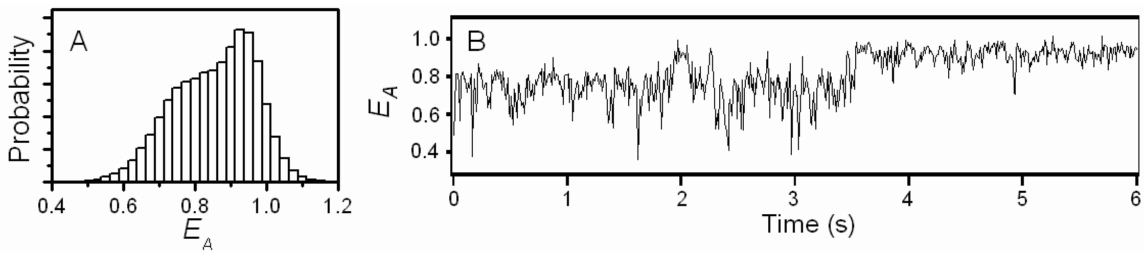


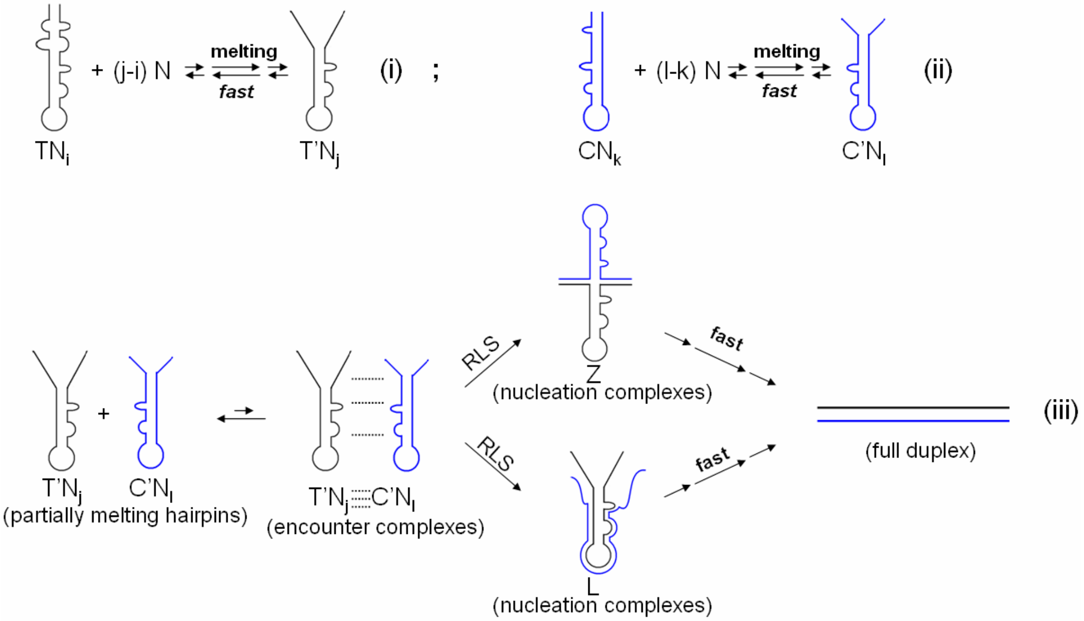
Figure 5.6: SM-FRET measurement of NC induced a partially melted second structure of TAR RNA.

The reaction was performed using immobilized Cy3/Cy5-TAR RNA at 0.2 mM $MgCl_2$ and 889 nM NC. (A) Ensemble E_A histograms by superimposing 80 single-molecule E_A histogram and (B) a single-molecule E_A time trajectory both show that the TAR RNA is switching between a closed and a partially melted structure. The E_A time trajectory shows that the time scale for the closed TAR RNA and semi-open TAR RNA is in seconds much faster than that of the C form/Y form interconverion of the TAR DNA (ca. in mili-seconds, see Chapter 3 for the detailed information)[67,130].

The mechanism portrayed in Scheme 5.3, assumes that rapid and reversible association and partial melting precedes a much slower nucleation event. The observation herein that the L1L2 loops of TAR are required (and in fact are sufficient) for rapid annealing - coupled with the previous observation that hairpins missing the L1L2 loop region do not undergo NC induced melting - strongly suggests that the encounter complex is formed by two hairpins with one or both of the hairpins in the "Y" conformation[67,130]. This conclusion is reflected in first step in equation (iii). The rate-limiting-step (RLS) for annealing, furthermore, is hypothesized to be local annealing at specific location along the hairpins, which can "nucleate" the annealing process. A steady-state solution of this mechanism predicts the second-order overall annealing kinetics (in non-aggregating conditions), which is in agreement with the experimental results. Furthermore, the close similarity of the annealing rate constant of TAR with cTAR, to that for other oligonucleotides that should strongly favor the zipper mechanism (i.e. -L3L4TAR and a short oligonucleotide targeted for L1L2 region[106]) strongly suggests that the zipper mechanism dominates the annealing reaction.

We have previously shown that a short, locally targeted oligonucleotide for the hairpin loop, including the L3L4 regions, is capable of annealing, although at a slower rate than the zipper nucleation route[106]. This implies that there may be two pathways for annealing for full length cTAR and TAR RNA. However, for the full length oligonucleotides, nucleation by either route ultimately leads to the same duplex annealed product. The similar rates of the loop and zipper routes (only an order of magnitude slower at low $MgCl_2$ concentrations) suggests that a common mechanism may be operating in all cases. Nucleation of annealing at two different locations corresponds to a "bifurcation" of the reaction path and technically corresponds to two distinct transition states (and mechanisms) for the reaction. It should be emphasized, however, that the two pathways are highly analogous, involving nucleation of annealing in locally single strand regions in the NC melted and associated complex of the two hairpins. Thus, a better description for the reaction than two-transition states may be a reaction mechanism in which the transition state is a broad multidimensional region along a "wide" reaction path with two or more, adjacent forms of the same type of transition states corresponding to different nucleation sites. Such complexities are not surprising in nucleic acid arrangements due to the rugged energy landscape [34,131,132].

The possibility that annealing can be initiated by nucleation through the hairpin loops (L4, HL region) has been recognized for some time, and has been previously assumed to occur through a kissing hairpin loop interaction mechanism, which is distinct from that summarized in Scheme 5.3[86,107,133]. The key feature of the kissing mechanism is an encounter complex that is stabilized by base pairing between the terminal loop region, i.e. the so-called kissing loop interaction. While the present set of experiments do not directly probe the possible involvement of kissing loop encounter complexes, the observations that TAR/-L3L4 nucleated annealing is as rapid as WT annealing indicates that kissing loop initiated annealing is not necessary for rapid annealing.



Scheme 5.3: The hypothetical kinetic scheme of TAR DNA annealing with its complements chaperoned by HIV-1 NC.

Here, T denotes TAR DNA, C denotes complementary cTAR DNA or TAR RNA, and N denotes NC. In this scheme, N bound to T and C leads to a partially melting structure, namely Y form of T (T') and C (C'). The subscripts, i, j, k, and l are used to describe the number of NC bound to nucleotides. Two partially melting hairpins form an encounter complex that leads to the formation of nucleation complexes. The annealing can go through either zipper nucleation or loop nucleation, therefore, forming zipper nucleation complexes (Z) or loop nucleation complexes (L) that leads to the formation of fully annealed duplexes.

The SM-FRET results strongly suggest that the common approach of investigating the NC induced annealing kinetics of nucleic acids with aggregates present must be undertaken with great caution due to the exceedingly complex and heterogeneous nature of these reactions at the molecular level. Nevertheless, the annealing reaction kinetics and mechanism within large scale aggregates is of interest due to its potential similarity with certain aspects of the annealing process in vivo. Indeed, future studies on annealing in large scale aggregates may shed light on how the spatial organization within complex NC/nucleic acid assemblies controls the rates and pathways of annealing in the reverse transcription mechanism of HIV-1.

5.4 CONCLUSIONS

By using a flow system with rapid mixing, we have been able to use a FRET-based approach to investigate the aggregation free annealing kinetics with full length, NC. Previously, in order to suppress aggregation effects on the annealing kinetics it was necessary to use a truncated NC(12-55) peptide that lacks the so-called aggregating terminal domain of NC[125]. New insights into the in vitro annealing mechanisms in the minus-strand transfer step of HIV-1 reverse transcription have been obtained. Specifically, the NC chaperoned irreversible annealing kinetics of a model TAR DNA hairpin sequence to the complementary cTAR (or TAR RNA) hairpin to form an extended duplex was investigated by SMS kinetic approaches. Specific information on the composition and secondary structures of key intermediates and transition state was obtained using single-molecule fluorescence tools (FRET, molecule counting, and correlation spectroscopy) in combination with a novel flow chamber approach involving rapid NC/nucleic acid mixing to substantially control aggregation. The results obtained from SMS demonstrate that the TAR hairpin reactant is predominantly a single NC coated hairpin with a dynamic secondary structure, involving an equilibrium between a “Y” shaped conformation and a closed one. The data further suggest that the nucleation of annealing occurs in an encounter complex that is formed by two hairpins.

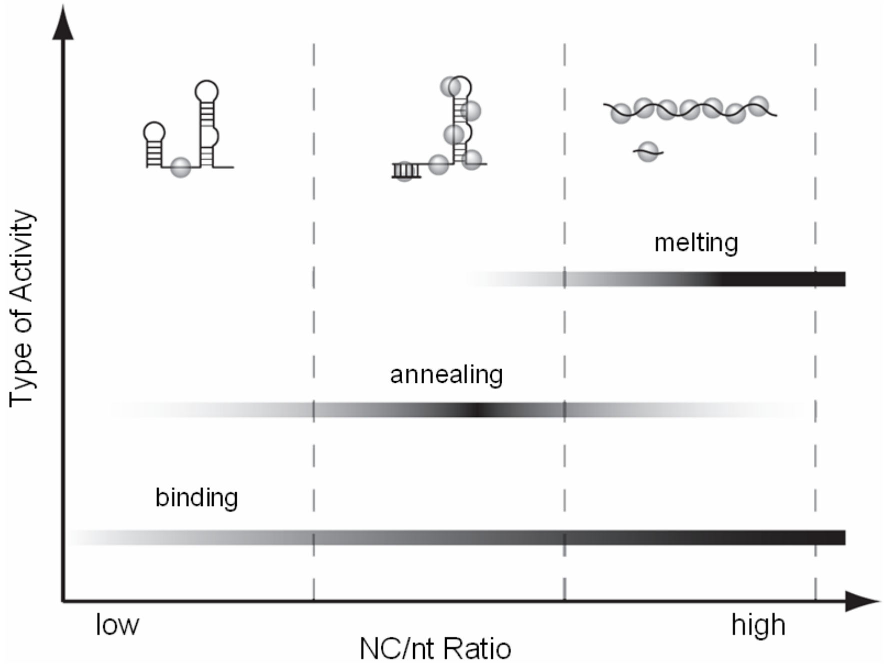
Chapter 6. Concentration-dependence of NC-Chaperoned Melting and Annealing of DNA Hairpin

6.1 INTRODUCTION

As a nucleic acid chaperone, NC promotes the annealing of complementary TAR DNA/RNA (cTAR) and meanwhile destabilizes the secondary structures in the L1L2 region of the TAR. Apparently, NC serves two distinct and counteracting roles on duplex stability of NAs, the stabilizing (or annealing) and destabilizing (or melting) two complementary strands. The duplex-stabilizing effect of NC is similar to the effect of increasing the ionic strength in the solution resulting from the screening the negative charges of the double-stranded NAs by basic amino acid residues and presumably through the N-terminal 3_{10} helix interacting non-specifically with phosphates[134-136]. Meanwhile, the duplex-destabilizing effect was observed when the NC concentrations were increased with the decreased salt concentrations. The addition of NC has shown to result in the destabilization of double-stranded NAs by 0.5-1.0 kcal per mol per base-pair[136,137].

Cristofari et. al proposed that the ability of NC on the stabilization and destabilization of two strands is dependent upon the NC/nt ratios[1]. At low NC/nt ratio, the NC enhances the annealing of two complementary strands whereas at a high NC/nt ratio, such enhancement does not necessarily take place due to the high amount of NC molecules bound to the NAs which causes the low likelihood of encounter of two complementary strands[138]. The resulting activities of NC at high NC/nt ratios are therefore melting rather than annealing two strands. Schemes 6.1 illustrates that the activities of NC based upon the NC/nt molar ratios can be divided into 3 types, binding, annealing and melting. In this model, the binding of NC at the NC/nt > 1:30 induces either the melting or annealing of two complementary strands presumably through the

NC-NC interactions depending upon the NC/nt ratios. In some cases, the NC-NC interactions between NC/NA complexes leads to the formation of large-scale-aggregates (LSA)[17]. Although the formation of LSA has been proposed to promote[18] as well as inhibit[139] the annealing of two complementary strands, the role of LSA in the annealing reaction is not well-understood.



Scheme 6.1: Degree of the NC binding versus NC's activity.

Left: "Binding mode": at NC/nt molar ratios of 1:100 to 1:40, all NC proteins bind to the NA, but do not exhibit any annealing activity, leading to the formation of complexes with one NC per NA. *Middle*: "Annealing mode": at NC/nt molar ratios of 1:30 to 1:10, all NC molecules are bound resulting in the formation of NC/NA complexes. It is believed that NC promotes the annealing of two strands by bringing NC/NA complexes together through the NC-NC interactions [1,140]. Optimal annealing activities are reported to be at ratios between 1:12 and 1:8[133,141]. *Right*: "Melting mode": at saturating NC conditions (i.e. NC/nt > 1:5), all the binding sites on NA molecules are occupied by NC molecules. The binding of NC stabilizes a relatively open conformation of NA as a consequence of the melting of inter- and intra-molecular secondary structures of NA[67,130]. In some cases, the individual NC-coated NA complexes serving as a fusion center lead to the formation of the large-scale aggregates(LSA)[17]. This figure is adapted from Cristofari et. al, 2002[1].

Studies on the NC's chaperone activities performed at various NC/nt ratios and salt conditions lead to a confusing conclusion about the binding properties of NC. For example, although preferential binding for singe-stranded NAs were reported by many groups [19-21,23-26], binding to double-stranded NAs has also been reported[25]. Many studies report relatively non-specific binding to single-stranded NA[25,32] but highly specific binding to UG- and TG-rich single stranded regions has also been reported[19-21]. GNG sequences have also been shown to be another preferential binding site[23,24].

The apparent discrepancies and complexities of NC's binding properties may be due to three main factors. First, LSA may have severely distorted binding studies in some publications. Second, the analysis of the binding results for double-stranded NAs may have been complicated in several cases by unanticipated NC-induced melting. Third, many of the studies on NC/NA interactions have been concerned with measuring the association constant for the NC/NA complex which is complicated by averaging over the multiple NC molecules bound to each NA molecule.

Thus, the apparent association constants are actually an ill-defined average over the individual binding sites. This can be a serious complication since only a small fraction of the total bound NC molecules and binding sites may play a role in NC's actual functions. Considerable progress has been made in understanding the distribution of NC binding sites along the NA strands[25]. These studies, however, generally have not distinguished between simple binding and functional effects such as NC-induced melting or annealing of NAs.

A new approach which combined the measurement of NC binding and NC-induced melting was applied using SM-FRET measurements. The degree of melting was measured using Cy3/Cy5-labeled DNA as a function of NC concentration. Two main parameters are derived for the secondary structure of a specific sequence by this technique. First, the dissociation constant for melting under the saturating NC concentrations is available by fitting the melting curve from the SM-FRET data with the suitable function. The other parameter is the number of the binding sites for NC on a specific NA. The dissociation constant for each binding sites will unravel if the binding of NC is cooperative.

6.2 DNA OLIGONUCLEOTIDES EMPLOYED AND THE EXPERIMENTAL METHOD

The simplest model by assuming that the biding of NC on NAs is cooperative can be obtained using Hill equation. Based on a 2-state model that a NA either has all or none of the binding sites occupied (all the binding sites are the same) so no intermediate states such as $NA{:}NC_1$ or $NA{:}NC_2$ are observed. The equation of this is expressed:

$$NA + nNC \rightleftarrows NA{:}NC_n \tag{6.1}$$

The dissociation constant for this reaction is

$$K = \frac{[NA][NC]^n}{[NA : NC_{\mathrm{n}}]} \tag{6.2}$$

and the fractional saturation due to all of the binding sites occupied is derived known as the Hill equation:

$$\%Saturation = \frac{[NC]^n}{K + [NC]^n} \tag{6.3}$$

where n is the average number of binding sites on a specific NA for NC. By fitting the melting curve obtained by SM-FRET, the dissociation constant as well as the number of the binding sites can be obtained. However, as mentioned above NC has been shown to have preferential binding to certain sequences and each binding site for NC might have non-equal dissociation constants. Such a 2-state-model based on the assumption that all the binding sites have equal affinity for NC is over-simplified. A modification on the Hill equation by considering each binding site with different affinity is derived (known as the Adair Equation, see Equation 6.6). The binding of NC to *i* different sites can be expressed in general:

$$NA{:}NC_{i-1} + NC \rightleftarrows NA{:}NC_i \tag{6.4}$$

The dissociation constant for binding the *i*-th site for NC is

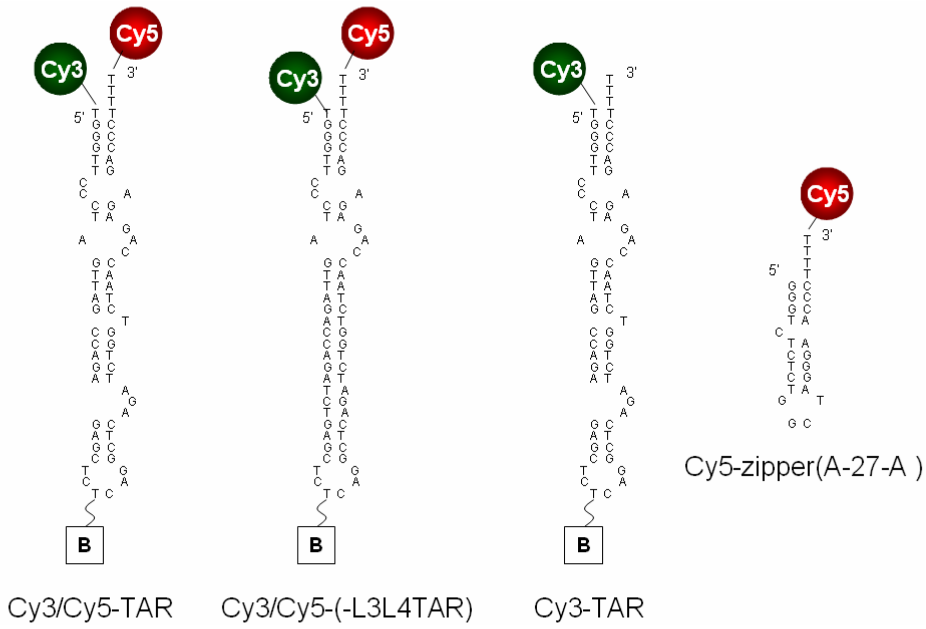
$$K_i = \frac{[NA:NC_{i-1}][NC]}{[NA:NC_i]} \tag{6.5}$$

In the case of the L1L2 region of TAR (expected to have up to 4 binding sites for NC molecules), the fraction of saturation of NC which is the fraction of occupied NC-binding sites divided by the total binding sites of NC can be expressed:

$$\%Saturation = \frac{\dfrac{[NC]}{4K_1} + \dfrac{[NC]^2}{2K_1K_2} + \dfrac{3[NC]^3}{4K_1K_2K_3} + \dfrac{[NC]^4}{K_1K_2K_3K_4}}{1 + \dfrac{[NC]}{K_1} + \dfrac{[NC]^2}{K_1K_2} + \dfrac{[NC]^3}{K_1K_2K_3} + \dfrac{[NC]^4}{K_1K_2K_3K_4}} \tag{6.6}$$

If $K_1 > K_2 > K_3 > K_4$, the binding of NC is positively cooperative (the binding of one NC promotes the binding of the next NC).

TAR DNA as well as –L3L4TAR were used to investigate the NC-induced melting of TAR at various NC concentrations. The zipper DNA was annealed with the TAR to obtain the information of NC-induced annealing at various NC concentrations (see Scheme 6.2 for the structures of oligonucleotides). In this measurement, instead of parking the scanning stage on individual molecules (as described in Chapter 3), the ensemble E_A histograms at each NC concentration were obtained using multiple scanning confocal images. The ensemble E_A histograms obtained by analyzing confocal images provided the E_A value from hundreds of molecules (as compared to c.a. 30 molecule for parking on individual molecules). Therefore, the data would not be biased by choosing a small amount of sampling. To achieve the SM-FRET measurement at various NC concentrations under a constant $MgCl_2$ concentration, one NC solution and two different concentrations of $MgCl_2$ solutions were applied. By adjusting the flow rates of 3 solutions, the final solution could be maintained at a constant 0.2 mM $MgCl_2$ and various NC concentrations.



Scheme 6.2: DNA oligonucleotides employed in this study.

6.3 RESULTS AND DISCUSSION

6.3.1 The Melting Curve of DNA Hairpin

Figure 6.1 demonstrated the melting curve of TAR obtained by the SM-FRET measurement. The data revealed that the transition of converting all the CTAR to the YTAR occurred around 178 nM. The ensemble E_A histograms obtained above and below the saturating NC concentrations indicated that the TAR was in the C form and Y form with a peak at $0.975E_A$ and $0.775E_A$, respectively. The % saturation (S%) obtained in the Figure 6.1 D can be calculated using the following equation:

$$< E_A > = E_{A,YTAR} S\% + E_{A,CTAR}(100 - S)\% \tag{6.7}$$

where $E_{A,YTAR}$=0.775 and $E_{A,CTAR}$ =0.975.

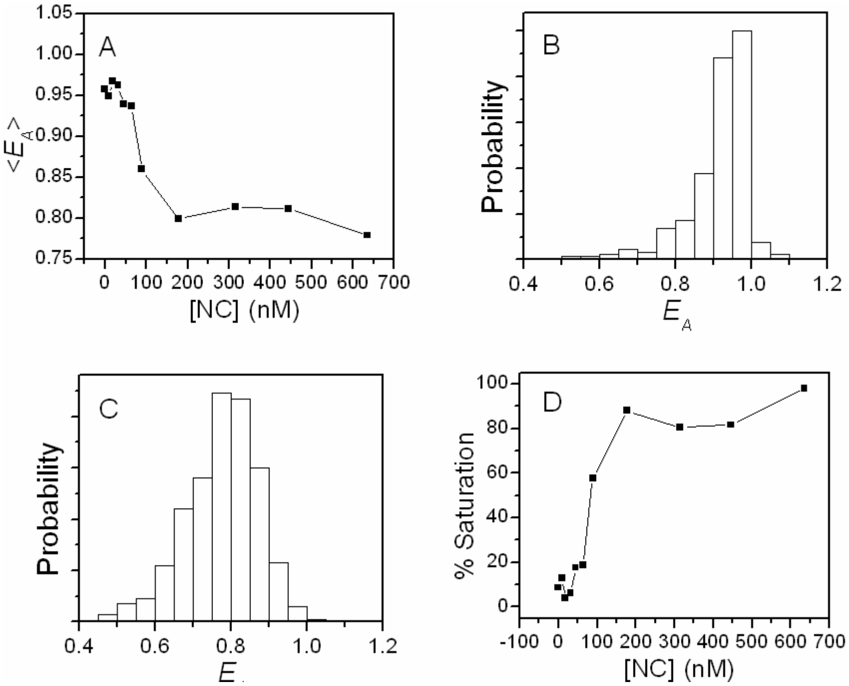


Figure 6.1: SM-FRET measurements on Cy3/Cy5-TAR at various NC concentrations. (A) The $<E_A>$ of TAR DNA as a function of NC concentrations. (B-C) The ensemble E_A histograms of TAR by superimposing E_A of all the molecules (B) below the saturating concentrations (0-63.5 nM) and (C) above the saturating concentrations (178-635 nM). (D) The percent saturation as a function of NC concentrations.

An attempt of using the 2-state-model to fit the data has been made. The fit to the % saturation with the Equation 6.3 (Hill Equation) yielded the dissociation constant, K with the error bar > 100%. The % saturation data was also fitted by the Adair equation (Equation 6.6). In this case, the fit to the data yielded that K_1 was 100 times bigger than K_2 while K_3 and K_4 was relatively similar. However, again the error bar for each K_i provided by the fitting was > 100%. The development of advanced fitting equation beyond the simple Hill or Adair Equation needs to be developed. Here, it is important to note that the $<E_A>$ obtained with the SM-FRET is susceptive to the photobleaching. Therefore, a threshold was applied to remove the molecules with $E_A < 0.5$, in all cases.

As mentioned earlier, there might be discrepancies among the binding sites of NC on TAR. NC might be tightly associated with some binding sites of the TAR as compared to others. In this case, the dissociation of NC with certain binding sites might be slower than others. To investigate this possibility, an additional step was adapted by flowing the buffer solutions in between various NC concentrations to wash out the bound NC. Figure 6.2 A shows that the $<E_A>$ obtained initially with the buffer only was 0.94. Once the binding of TAR with NC (at 635 nM) was saturated, the E_A went down to 0.73. However, flowing the sample with the buffer for 15 min did not bring the $<E_A>$ all the way up to 0.94 (data point 1 in Figure 6.2 A). To further investigate that the possibility of tight binding sites causes the lower $<E_A>$ (0.873±0.009), the reversible melting was also performed with the –L3L4TAR by incubating the –L3L4TAR under a sequential flowing solutions (see Figure 6.1), first the buffer, then the saturating NC (445 nM) and the buffer again for at least 30 min for each condition. The data with –L3L4TAR shows that the $<E_A>$ of –L3L4TAR at two buffer periods exhibited the similar value (0.967 ± 0.003 in the period 1 and 0.954 ± 0.005 in the period 3). It has been shown that NC shifted the equilibrium of TAR/-L3L4TAR from the C form to the Y form relying on partially melting the L1L2 region [67,130]. It is reasonable to assume that the binding sites for NC on the L1L2 region of TAR and –L3L4TAR is the same. Therefore, it is unlikely that there exists an intrinsic tight binding site on L1L2 region of TAR but not that of –L3L4TAR.

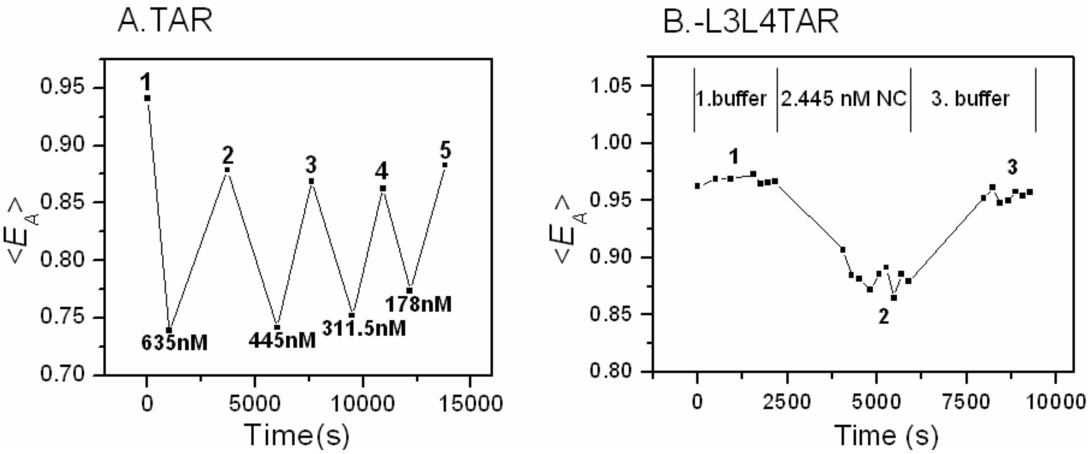


Figure 6.2: $<E_A>$ of TAR/-L3L4TAR at different times with various NC concentrations.

(A) The $< E_A >$ of TAR DNA at different times. The measurement started from high to low NC concentrations. Between each NC concentration, the sample was flow with the buffer only (point 1-4) to wash out the bound NC for at least 15 min. (B) The $< E_A >$ of –L3L4TAR DNA at different times. The condition for the period 1,2 and 3 was buffer only, 445 nM NC and buffer only again, respectively.

The other possibility for the lower $<E_A>$ for TAR during the NC unbinding is that the rapid removal of NC due to the change of the flow condition causes the unwanted misfolded states. Single-molecule force-jump and force-ramp experiments have shown that under the condition of rapidly refolding the TAR RNA, the refolding of the TAR RNA exhibits a multiple-step[142]. This suggests the misfolded state as intermediates during the folding of TAR RNA; however, the misfolded states are observed with RNA more often than DNA. In addition, the mfold predicted a stable alternative folded state of the TAR with the activation energy similar to that of the CTAR. Such a folded state has a close end-to-end distance between two termini which might produce a sub-population hidden in the CTAR or YTAR distribution. Nevertheless, this misfolded structure involves the rearrangements of the L3L4 region which might be difficult to form due to the kinetic barrier required to unfold the L3L4 region. Therefore, the idea of the misfolded states of TAR DNA needed to be further investigated.

6.3.2 NC's Ability on Melting versus Annealing Nucleic Acids

The direct SM-FRET on Cy3/Cy5-labeled TAR for measuring local melting can detect the melting in L1L2 regions. However, the difference between the E_A of YTAR and CTAR is limited. Data from a more sensitive, but indirect approach for NC-induced melting is shown in Fig. 6.3. In this technique, the yield of NC catalyzed annealing of TAR DNA hairpin with a short targeted oligomer, zipper DNA (see Scheme 6.2) is measured by SM-FRET as a function of the NC concentrations. This measurement leads to determining the annealing kinetics as a function of NC concentration. The annealing of TAR with the zipper DNA turns on at c.a. 80 nM NC (shown in Figure 6.3 A), and the % annealed reaches the plateau at around 160 nM(shown in Figure 6.3 B).

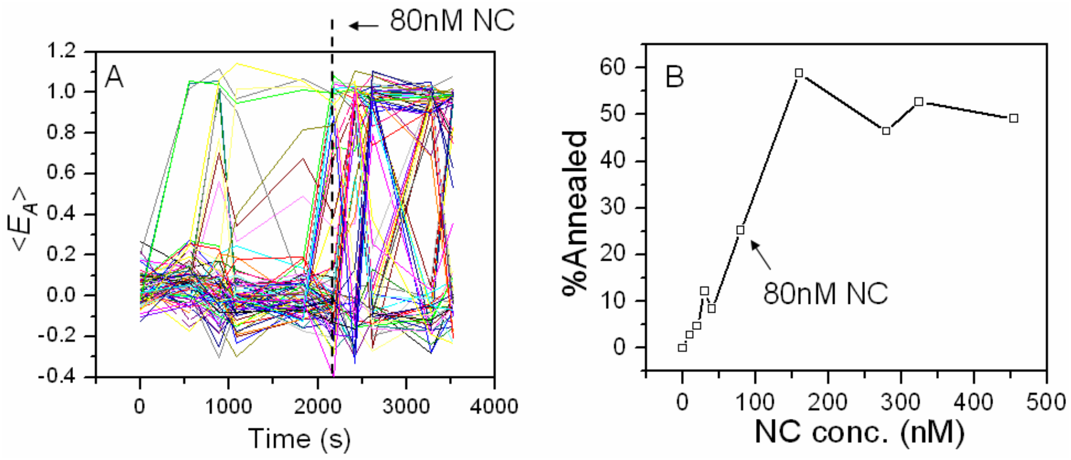


Figure 6.3: <E_A> of TAR annealing with 20 nM zipper DNA at different times with various NC concentrations.

(A) The < E_A > of TAR DNA at different times. (B) The % annealed of TAR at different NC concentrations.

In turn, the kinetic data can be used to infer the degree of melting of the TAR as a function of NC concentration. However, a model is needed to be developed to correlate the degree of the melting versus the equilibrium constant of the annealing reaction.

6.4 CONCLUSIONS

Studies on various oligonucleotides at different NC/nt ratios and salt conditions led to a confusing picture of NC's binding properties presumably due to the binding of NC-induced different degree of melting or annealing of NAs. The first measurement of correlating the binding of NC to the melting/annealing of NAs was performed using SM-FRET measurements. The data shows that the melting of TAR is not a simple 2-state process and perhaps involves an additional alternative folded state. The annealing of the TAR with a short oligomer as a function of NC concentrations can be used to infer the degree of melting of the TAR. However, a suitable model will need to be developed in the future in order to quantitatively analyze the data.

Chapter 7. General Conclusions

Using time-resolved SM-FRET, the conformational distribution and dynamics of TAR DNA has been obtained in the presence and absence of WT HIV-1 NC[67,130]. Various WT and mutant TAR DNA constructs with different numbers of internal bulges deleted were examined. The SM-FRET date indicates that NC shifts the structure of TAR from a predominantly closed structure (CTAR) towards a partially open structure of the TAR (YTAR). NC induces the YTAR structure mainly relying on the interaction with two terminal bulges/loops (L1L2) regions of the TAR. The predominantly YTAR has been suggested to be a key intermediate in NC-facilitated TAR RNA/DNA annealing[139].

Various complementary DNA oligonucleotides designed to mimicking the initial annealing step (nucleation) were employed to anneal with the TAR in the presence of NC. In these studies, two annealing pathways, “zipper” and “kissing” pathway were probed using the SM-FRET annealing kinetic measurements. In the zipper pathway, the annealing of the TAR initiates through the 3’/5’ termini of the TAR while in the kissing pathway, the annealing of the TAR initiates through the hairpin loop region of the TAR. Although the formation of multiple intermediates i.e. zipper and kissing intermediates was observed, the annealing kinetics through the zipper route occurs faster than the zipper route. These studies suggest that multiple annealing pathways can occur for the NC-mediated TAR RNA/DNA annealing step of minus-strand transfer[106].

Using a combination of SMS techniques, including FRET, molecule counting and correlation spectroscopy, more mechanistic insights into NC-mediated TAR RNA/DNA annealing have been obtained[139]. In this study, annealing kinetics of NC-mediated full-length TAR DNA to TAR RNA or various complementary TAR (cTAR) DNA constructs with different numbers of internal bulges deleted were obtained using SM-FRET measurement combined with a flow system containing rapid mixing. The ability of this approach is to control NC-induced aggregation and therefore observe the annealing kinetics without the influence of the large-scale aggregation. Under this condition, the data shows that the structure of the TAR DNA is predominantly a single NC-coated hairpin with a dynamic partially YTAR conformation. The two terminal loops/bulges is critical not only for NC-induced melting of TAR DNA[139] but also for NC-facilitated rapid annealing of TAR and TAR RNA (or cTAR). The annealing kinetic data indicate that the annealing rate constant of the WT TAR DNA hairpin with TAR RNA (or cTAR) hairpin is similar to that of the zipper mimic oligomers, which suggests that the zipper mechanism also dominates the TAR DNA/RNA annealing reaction. Moreover, the nucleation intermediate of the annealing reaction is formed by two hairpins, with one or both of the hairpins in the "Y" conformation.

References

1. Cristofari, G.; Darlix, J.-L., The ubiquitous nature of RNA chaperone proteins. *Prog. Nucleic Acid Res. Mol. Biol.* **2002,** 72, 223-268.
2. Ho, D. D.; Neumann, A. U.; Perelson, A. S.; Chen, W.; Leonard, J. M.; Markowitz, M., Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* **1995,** 373, (6510), 123-126.
3. Wei, X.; Ghosh, S. K.; Taylor, M. E.; Johnson, V. A.; Emini, E. A.; Deutsch, P.; Lifson, J. D.; Bonhoeffer, S.; Nowak, M. A.; Hahn, B. H.; Saag, M. S.; Shaw, G. M., Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* **1995**, 373, (6510), 117-122.
4. Coffin, J. M.; Hughes, S. H.; Varmus, H. E., Retroviruses. *Cold Spring Harbor Laboratory Press*, **1997**.
5. Hu, W. S.; Temin, H. M., Retroviral recombination and reverse transcription. *Science* **1990,** 250, (4985), 1227-33.
6. Pullen, K. A.; Champoux, J. J., Plus-strand origin for human immunodeficiency virus type 1: implications for integration. *J. Virol.* **1990,** 64, (12), 6274–6277.
7. Resnick, R.; Omer, C. A.; Faras, A. J., Involvement of retrovirus reverse transcriptase-associated RNase H in the initiation of strong-stop (+) DNA synthesis and the generation of the long terminal repeat. *J. Virol.* **1984,** 51, (3), 813-821.
8. Berg, J. M., Potential metal-binding domains in nucleic acid binding proteins. *Science* **1986,** 232, (4749), 485-487.
9. Covey, S. N., Amino acid sequence homology in *gag* region of reverse transcribing elements and the coat protein gene of cauliflower mosiac virus. *Nucleic Acids Res.* **1986,** 14, (2), 623-633.
10. Green, L. M.; Berg, J. M., Retroviral nucleocapsid protein-metal ion interactions: folding and sequence variants. *Proc. Natl. Acad. Sci. USA.* **1990,** 87, (16), 6403-6407.
11. Henderson, L. E.; Copeland, T. D.; Sowder, R. C.; Smythers, G. W.; Oroszlan, S., Primary structure of the low molecular weight nucleic acid-binding proteins of murine leukemia viruses. *J. Biol. Chem.* **1981,** 256, (16), 8400-8406.
12. Darlix, J.-L.; Lapadat-Tapolsky, M.; de Rocquigny, H.; Roques, B. P., First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses. *J. Mol. Biol.* **1995,** 254, (4), 523-537.
13. Turner, B. G.; Summers, M. F., Structural biology of HIV. *J. Mol. Biol.* **1999,** 285, 1-32.
14. Tanchou, V.; Gabus, C.; Rogemond, V.; Darlix, J.-L., Formation of stable and functional HIV-1 nucleoprotein complexes *in vitro*. *J. Mol. Biol.* **1995,** 252, (5), 563-571.
15. Lee, B. M.; De Guzman, R. N.; Turner, B. G.; Tjandra, N.; Summers, M. F., Dynamical behavior of the HIV-1 nucleocapsid protein. *J. Mol. Biol.* **1998,** 279, (3), 633-649.

16. Urbaneja, M. A.; Kane, B. P.; Johnson, D. G.; Gorelick, R. J.; Henderson, L. E.; Casas-Finet, J. R., Binding properties of the human immunodeficiency virus type 1 nucleocapsid protein p7 to a model RNA: elucidation of the structural determinants for function. *J. Mol. Biol.* **1999,** 287, (1), 59-75.
17. Stoylov, S. P.; Vuilleumier, C.; Stoylova, E.; De Rocquigny, H.; Roques, B. P.; Gerard, D.; Mely, Y., Ordered aggregation of ribonucleic acids by the human immunodeficiency virus type 1 nucleocapsid protein. *Biopolymers* **1997,** 41, (3), 301-312.
18. Le Cam, E.; Coulaud, D.; Delain, E.; Petitjean, P.; Roques, B. P.; Gerard, D.; Stoylova, E.; Vuilleumier, C.; Stoylov, S. P.; Mely, Y., Properties and growth mechanism of the ordered aggregation of a model RNA by the HIV-1 nucleocapsid protein: an electron microscopy investigation. *Biopolymers* **1998,** 45, (3), 217-229.
19. Fisher, R. J.; Rein, A.; Fivash, M.; Urbaneja, M. A.; Casas-Finet, J. R.; Medaglia, M.; Henderson, L. E., Sequence-specific binding of human immunodeficiency virus type 1 nucleocapsid protein to short oligonucleotides. *J. of Virol.* **1998,** 72, (3), 1902-1909.
20. Berglund, J. A.; Charpentier, B.; Rosbash, M., A high affinity binding site for the HIV-1 nucleocapsid protein. *Nucleic Acids Res.* **1997,** 25, (5), 1042-1049.
21. D'Souza, V.; Summers, M. F., Structural basis for packaging the dimeric genome of Moloney murine leukemia virus. *Nature* **2004,** 431, (7008), 586-590.
22. Williams, M. C.; Rouzina, I.; Wenner, J. R.; Gorelick, R. J.; Musier-Forsyth, K.; Bloomfield, V. A., Mechanism for nucleic acid chaperone activity of HIV-1 nucleocapsid protein revealed by single molecule stretching. *Proc. Natl. Acad. Sci. USA* **2001,** 98, (11), 6121-6126.
23. Amarasinghe, G. K.; de Guzman, R. N.; Turner, R. B.; Chancellor, K. J.; Wu, Z. R.; Summers, M. F., NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the Ψ-RNA packaging signal. Implications for genome recognition. *J. Mole. Biol.* **2000,** 301, 491-511.
24. De Guzman, R. N.; Wu, Z. R.; Stalling, C. C.; Pappalardo, L.; Borer, P. N.; Summers, M. F., Structure of the HIV-1 nucleocapsid protein bound to the SL3 Ψ-RNA recognition element. *Science* **1998,** 279, 384-388.
25. Urbaneja, M. A.; Wu, M.; Casas-Finet, J. R.; Karpel, R. L., HIV-1 nucleocapsid protein as a nucleic acid chaperone: spectroscopic study of its helix-destabilizing properties, structural binding specificity, and annealing activity. *J. Mole. Biol.* **2002,** 318, (3), 749-764.
26. Vuilleumier, C.; Bombarda, E.; Morellet, N.; Gerard, D.; Roques, B. P.; Mely, Y., Nucleic Acid Sequence Discrimination by the HIV-1 Nucleocapsid Protein NCp7: A Fluorescence Study. *Biochemistry* **1999,** 38, (51), 16816-16825.
27. Henderson, L. E.; Bowers, M. A.; Sowder, R. C. I.; Serabyn, S. A.; Johnson, D. G.; Bess, J. W. J.; Arthur, L. O.; Bryant, D. K.; Fenselau, C., Gag proteins of the highly replicative MN strain of human immunodeficiency virus type 1: posttranslational modifications, proteolytic processings, and complete amino acid sequences. *J. Virol.* **1992,** 66, 1856-1865.

28. Khan, R.; Giedroc, D. P., Nucleic acid binding properties of recombinant Zn^{2+} HIV-1 nucleocapsid protein are modulated by COOH-terminal processing. *J. Biol. Chem.* **1994,** 269, (36), 22538-22546.
29. Karpel, R. L.; Henderson, L. E.; Oroszlan, S., Interactions of retroviral structural proteins with single-stranded nucleic acids. *J. Biol. Chem.* **1987,** 262, (11), 4961-4967.
30. Whiting, S. H.; Champoux, J. J., Strand displacement synthesis capability of Moloney murine leukemia virus reverse transcriptase. *J. Virol.* **1994,** 68, (8), 4747-4758.
31. Dibhajj, F.; Khan, R.; Giedroc, D. P., Retroviral Nucleocapsid Proteins Posses Potent Nucleic-Acid Strand Renaturation Activity. *Protein Sci.* **1993,** 2, (2), 231-243.
32. Mély, Y.; de Rocquigny, H.; Sorinas-Jimeno, M.; Keith, G.; Roques, B. P.; Marquet, R.; Gerard, D., Binding of the HIV-1 nucleocapsid protein to the primer $tRNA_3^{Lys}$ *in vitro*, is essentially not specific. *J. Biol. Chem.* **1995,** 270, (4), 1650-1656.
33. Shubsda, M. F.; Paoletti, A. C.; Hudson, B. S.; Borer, P. N., Affinities of Packaging Domain Loops in HIV-1 RNA for the Nucleocapsid Protein. *Biochemistry* **2002,** 41, (16), 5276-5282.
34. Herschlag, D., RNA chaperones and the RNA folding problem. *J. Biol. Chem.* **1995,** 270, (36), 20871-20874.
35. Rein, A.; Henderson, L. E.; Levin, J. G., Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication. *Trends Biochem. Sci.* **1998,** 23, (8), 297-301.
36. Schroeder, R.; Barta, A.; Semrad, K., Strategies for RNA folding and assembly. *Nat. Rev. Mol. Cell Biol.* **2004,** 5, (11), 908-919.
37. Tsuchihashi, Z.; Brown, P. O., DNA Strand Exchange and Selective DNA Annealing Promoted by the Human Immunodeficiency Virus Type 1 Nucleocapsid Protein. *J. Virol.* **1994,** 68, (9), 5863-5870.
38. Levin, J. G.; Guo, J.; Rouzina, I.; Musier-Forsyth, K., Nucleic Acid Chaperone Activity of HIV-1 Nucleocapsid Protein: Critical Role in Reverse Transcription and Molecular Mechanism. *Prog. Nucleic Acid Res. Mol. Biol.* **2005,** 80, 217-286.
39. Lapadat-Tapolsky, M.; De Rocquigny, H.; Van Gent, D.; Roques, B.; Plasterk, R.; Darlix, J. L., Interactions between HIV-1 nucleocapsid protein and viral DNA may have important functions in the viral life cycle. *Nucleic Acids Res.* **1993,** 21, (4), 831-839.
40. Aldovini, A.; Young, R. A., Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. *J. Virol.* **1990,** 64, 1920-1926.
41. Dorfman, T.; Luban, J.; Goff, S. P.; Haseltine, W. A.; Göttlinger, H. G., Mapping of functionally important residues of a cysteine-histidine box in the human immunodeficiency virus type 1 nucleocapsid protein. *J. Virol.* **1993,** 67, (10), 6159-6169.
42. Berkowitz, R.; Fisher, J.; Goff, S. P., RNA packaging. *Curr. Top Microbiol. Immunol.* **1996,** 214, 177-218.

43. Freed, E. O., HIV-1 Gag Proteins: Diverse Functions in the Virus Life Cycle. *Virology* **1998,** 251, (1), 1-15.
44. De Rocquigny, H.; Gabus, C.; Vincent, A.; Fournié-Zaluski, M.-C.; Roques, B.; Darlix, J.-L., Viral RNA annealing activities of human immunodeficiency virus type 1 nucleocapsid protein require only peptide domains outside the zinc fingers. *Proc. Natl. Acad. Sci. USA* **1992,** 89, 6472-6476.
45. Isel, C.; Westhof, E.; Massire, C.; Le Grice, S. F. J.; Ehresmann, B.; Ehresmann, C.; Marquet, R., Structural basis for the specificity of the initiation of HIV-1 reverse transcription. *EMBO J.* **1999,** 18, (4), 1038-1048.
46. Feng, Y. X.; Campbell, S.; Harvin, D.; Ehresmann, B.; Ehresmann, C.; Rein, A., The human immunodeficiency virus type 1 Gag polyprotein has nucleic acid chaperone activity: possible role in dimerization of genomic RNA and placement of tRNA on the primer binding site. *J. Virol.* **1999,** 73, (5), 4251-4256.
47. De Rocquigny, H.; Gabus, C.; Vincent, A.; Fournié-Zaluski, M. C.; Roques, B.; Darlix, J. L., Viral RNA annealing activities of human immunodeficiency virus type 1 nucleocapsid protein require only peptide domains outside the zinc fingers. *Proc. Natl. Acad. Sci. U. S. A.* **1992,** 89, (14), 6472-6476.
48. Wu, T.; Guo, J.; Bess, J.; Henderson, L. E.; Levin, J. G., Molecular requirements for human immunodeficiency virus type 1 plus-strand transfer: analysis in reconstituted and endogenous reverse transcription systems. *J. Virol.* **1999,** 73, (6), 4794-4805.
49. Guo, J.; Wu, T.; Anderson, J.; Kane, B. F.; Johnson, D. G.; Gorelick, R. J.; Henderson, L. E.; Levin, J. G., Zinc finger structures in the human immunodeficiency virus type 1 nucleocapsid protein facilitate efficient minus- and plus-strand transfer. *J. Virol.* **2000,** 74, (19), 8980-8988.
50. Johnson, P. E.; Turner, R. B.; Wu, Z. R.; Hairston, L.; Guo, J.; Levin, J. G.; Summers, M. F., A mechanism for plus-strand transfer enhancement by the HIV-1 nucleocapsid protein during reverse transcription. *Biochemistry* **2000,** 39, (31), 9084-9091.
51. Muthuswami, R.; Chen, J.; Burnett, B. P.; Thimmig, R. L.; Janjic, N.; McHenry, C. S., The HIV plus-strand transfer reaction: determination of replication-competent intermediates and identification of a novel lentiviral element, the primer over-extension sequence. *J. Mol. Biol.* **2002,** 315, (3), 311-323.
52. Barat, C.; Schatz, O.; Le Grice, S.; Darlix, J.-L., Analysis of the Interactions of HIV1 Replication Primer tRNA^{Lys,3} with Nucleocapsid Protein and Reverse Transcriptase. *J. Mol. Biol.* **1993,** 231, (2), 185-190.
53. Franke, E. K.; Yuan, H. E.; Bossolt, K. L.; Goff, S. P.; Luban, J., Specificity and sequence requirements for interactions between various retroviral Gag proteins. *J. Virol.* **1994,** 68, 5300-5305.
54. Zhang, Y.; Barklis, E., Effects of nucleocapsid mutations on human immunodeficiency virus assembly and RNA encapsidation. *J. Virol.* **1997,** 71, 6765-6776.
55. Zhang, Y.; Qian, H.; Love, Z.; Barklis, E., Analysis of the assembly function of the human immunodeficiency virus type 1 Gag protein nucleocapsid domain. *J. Virol.* **1998,** 72, (3), 1782-1789.

56. Wu, W.; Henderson, L. E.; Copeland, T. D.; Gorelick, R. J.; Bosche, W. J.; Rein, A.; Levin, J. G., Human immunodeficiency virus type 1 nucleocapsid protein reduces reverse transcriptase pausing at a secondary structure near the murine leukemia virus polypurine tract. *J. Virol.* **1996,** 70, 7132-7142.

57. Chan, B.; Weidemaier, K.; Yip, W. T.; Barbara, P. F.; Musier-Forsyth, K., Intra-tRNA distance measurements for nucleocapsid protein dependent tRNA unwinding during priming of HIV reverse transcription. *Proc. Natl. Acad. Sci. USA* **1999,** 96, 459.

58. Wu, T.; Guo, J.; Bess, J.; Henderson, L. E.; Levin, J. G., Molecular Requirements for Human Immunodeficiency Virus Type 1 Plus-Strand Transfer: Analysis in Reconstituted and Endogenous Reverse Transcription Systems. *J. Virol.* **1999,** 73, 4794-4805.

59. Guo, J.; Wu, T.; Anderson, J.; Kane, B. F.; Johnson, D. G.; Gorelick, R. J.; Henderson, L. E.; Levin, J. G., Zinc finger structures in the human immunodeficiency virus type 1 nucleocapsid protein facilitate effecient minus- and plus-strand transfer. *J. Virol.* **2000,** 74, (19), 8980-8988.

60. Johnson, P. E.; Turner, R. B.; Wu, Z. R.; Hairston, L.; Guo, J.; Levin, J. G.; Summers, M. F., A mechanism for plus-strand transfer enhancement by the HIV-1 nucleocapsid protein during reverse transcription. *Biochemistry* **2000,** 39, 9084-9091.

61. You, J. C.; McHenry, C. S., Human immunodeficiency virus nucleocapsid protein accelerates strand transfer of the terminally redundant sequences involved in reverse transcription. *J. Biol. Chem.* **1994,** 269, (50), 31491-31495.

62. Darlix, J.-L.; Vincent, A.; Gabus, C.; de Rocquigny, H.; Roques, B., Trans-activation of the 5' to 3' viral DNA strand transfer by nucleocapsid protein during reverse transcription of HIV-1 RNA. *C. R. Acad. Sci., Paris Life Sciences* **1993,** 316, (8), 763-771.

63. Allain, B.; Lapadat-Tapolsky, M.; Berlioz, C.; Darlix, J.-L., Transactivation of the Minus-strand Transfer by Nucleocapsid Protein during Reverse Transcription of the Retroviral Genome. *EMBO J.* **1994,** 13, (4), 973-981.

64. Peliska, J. A.; Balasubramanian, S.; Giedroc, D.; Benkovic, S. J., Recombinant HIV-1 Nucleocapsid Protein Accelerates HIV-1 Reverse Transcriptase Catalyzed DNA Strand Transfer Reactions and Modulates RNase H Activity. *Biochemistry* **1994,** 33, 13817-13823.

65. Kim, J. K.; Palaniappan, C.; Wu, W.; Fay, P. J.; Bambara, R. A., Evidence for a unique mechanism of strand transfer from the transactivation response region of HIV-1. *J. Biol. Chem.* **1997,** 272, 16769-16777.

66. Guo, J.; Henderson, L. E.; Bess, J.; Kane, B.; Levin, J. G., Human immunodeficiency virus type 1 nucleocapsid protein promotes efficient strand transfer and specific viral DNA synthesis by inhibiting TAR-dependent self-priming from minus-strand strong-stop DNA. *J. Virol.* **1997,** 71, (7), 5178-5188.

67. Cosa, G.; Harbron, E. J.; Zeng, Y.; Liu, H.-W.; O'Connor, D. B.; Eta-Hosokawa, C.; Musier-Forsyth, K.; Barbara, P. F., Secondary Structure and Secondary Structure Dynamics of DNA Hairpins Complexed with HIV-1 NC Protein. *Biophys. J.* **2004,** 87, (4), 2759-2767.

68. Azoulay, J.; Clamme, J.-P.; Darlix, J.-L.; Roques, B. P.; Mely, Y., Destabilization of the HIV-1 Complementary Sequence of TAR by the Nucleocapsid Protein Through Activation of Conformational Fluctuations. *J. Mol. Biol.* **2003,** 326, (3), 691-700.
69. Beltz, H.; Azoulay, J.; Bernacchi, S.; Clamme, J.-P.; Ficheux, D.; Roques, B.; Darlix, J.-L.; Mely, Y., Impact of the Terminal Bulges of HIV-1 cTAR DNA on its Stability and the Destabilizing Activity of the Nucleocapsid Protein NCp7. *J. Mole. Biol.* **2003,** 328, (1), 95-108.
70. Beltz, H.; Piemont, E.; Schaub, E.; Ficheux, D.; Roques, B.; Darlix, J. L.; Mély, Y., Role of the structure of the top half of HIV-1 cTAR DNA on the nucleic acid destabilizing activity of the nucleocapsid protein NCp7. *J. Mol. Biol.* **2004,** 338, (4), 711-723.
71. Bernacchi, S.; Stoylov, S.; Piemont, E.; Ficheux, D.; Roques, B. P.; Darlix, J.-L.; Mely, Y., HIV-1 Nucleocapsid Protein Activates Transient Melting of Least Stable Parts of the Secondary Structure of TAR and its Complementary Sequence. *J. Mol. Biol.* **2002,** 317, 385-399.
72. Lapadat-Tapolsky, M.; Gabus, C.; Rau, M.; Darlix, J.-L., Possible Roles of HIV-1 Nucleocapsid Protein in the Specificity of Proviral DNA Synthesis and in Its Variability. *J. Mol. Biol.* **1997,** 268, (2), 250-260.
73. Driscoll, M. D.; Hughes, S. H., Human immunodeficiency virus type 1 nucleocapsid protein can prevent self-priming of minus-strand strong stop DNA by promoting the annealing of short oligonucleotides to hairpin sequences. *J. Virol.* **2000,** 74, (19), 8785-8792.
74. Ha, T.; Enderle, T.; Chemla, D. S.; Selvin, P. R.; Weiss, S., Single Molecule Dynamics Studied by Polarization Modulation. *Phys. Rev. Lett.* **1996,** 77, (19), 3979-3982.
75. Magde, D.; Elson, E.; Webb, W. W., Thermodynamic fluctations in a reacting system. Measurement by fluorescence correlation spectroscopy. *Phys. Rev. Lett.* **1972,** 29, (11), 705-708.
76. Stryer, L., Fluorescence energy transfer as a spectroscopic ruler. *Ann. Rev. Biochem.* **1978,** 47, 819-846.
77. Clegg, R. M., *Fluorescence resonance energy transfer*. Wiley: New York, 1996; p 179–252.
78. Ha, T.; Rasnik, I.; Cheng, W.; Babcock, H. P.; Gauss, G.; Lohman, T. M.; Chu, S., Initiation and reinitiation of DNA unwinding by the E. Coli Rep helicase. *Nature* **2002,** 419, 638-641.
79. Rasnik, I.; Myong, S.; Cheng, W.; Lohman, T. M.; Ha, T., DNA-binding Orientation and Domain Conformation of the E. coli Rep Helicase Monomer Bound to a Partial Duplex Junction: Single-molecule Studies of Fluorescently Labeled Enzymes. *J. Mol. Biol.* **2004,** 336, (2), 395-408.
80. Rasnik, I.; Myong, S.; Ha, T., Unraveling helicase mechanisms one molecule at a time *Nucleic Acids Res.* **2006,** 34, 4225 - 4231.
81. Hess, S. T.; Huang, S.; Heikal, A. A.; Webb, W. W., Biological and Chemical Applications of Flurescence Correlation Spectroscopy: A Review. *Biochemistry* **2002,** 41, (3), 697-705.

82. Lakowicz, J. R., Energy Transfer. In *Principles of Fluorescence Spectroscopy*, Kluwer Academic: New York, 1999; 368-391.
83. Krichevsky, O.; Bonnet, G., Fluorescence correlation spectroscopy: the technique and its applications. *Rep. Prog. Phys.* **2002,** 65, (2), 251-297.
84. Kim, H. D.; Nienhaus, G. U.; Ha, T.; Orr, J. W.; Williamson, J. R.; Chu, S., Mg^{2+}-dependent conformational change of RNA studied by fluorescence correlation and FRET on immobilized single molecules. *Proc. Natl. Acad. Sci. USA* **2002,** 99, 4284-4289.
85. Berkhout, B.; Vastenhouw, N. L.; Klasens, B. I. F.; Huthoff, H., Structural features in the HIV-1 repeat region facilitate strand transfer during reverse transcription. *RNA* **2001,** 7, (8), 1097-114.
86. Vo, M.-N.; Barany, G.; Rouzina, I.; Musier-Forsyth, K., Mechanistic Studies of Mini-TAR RNA/DNA Annealing in the Absence and Presence of HIV-1 Nucleocapsid Protein. *J. Mole. Biol.* **2006,** 363, (1), 244-261.
87. Gorelick, R. J.; Nigida, S. M., Jr.; Bess, J. W., Jr.; Arthur, L. O.; Henderson, L. E.; Rein, A., Noninfectious human immunodeficiency virus type 1 mutants deficient in genomic RNA. *J. Virol.* **1990,** 64, (7), 3207-3211.
88. Zhang, Y.; Barklis, E., Nucleocapsid protein effects on the specificity of retrovirus RNA encapsidation. *J. Virol.* **1995,** 69, (9), 5716-5722.
89. Schwartz, M. D.; Fiore, D.; Panganiban, A. T., Distinct functions and requirements for the cys-his boxes of the human immunodeficiency virus type 1 nucleocapsid protein during RNA encapsidation and replication. *J. Virol.* **1997,** 71, 9295-9350.
90. Schmalzbauer, E.; Strack, B.; Dannull, J.; Guehmann, S.; Moelling, K., Mutations of basic amino acids of NCp7 of human immunodeficiency virus type 1 affect RNA binding in vitro. *J. Virol.* **1996,** 70, (2), 771-777.
91. Poon, D. T.; Wu, J.; Aldovini, A., Charged amino acid residues of human immunodeficiency virus type 1 nucleocapsid p7 protein involved in RNA packaging and infectivity. *J. Virol.* **1996,** 70, (10), 6607-6616.
92. Berg, J. M.; Shi, Y., The galvanization of biology: a growing appreciation for the roles of zinc. *Science* **1996,** 271, (5252), 1081-1085.
93. Rice, W. G.; Supko, J. G.; Malspeis, L.; Buckheit, R. W., Jr.; Clanton, D.; Bu, M.; Graham, L.; Schaeffer, C. A.; Turpin, J. A.; Domagala, J.; et al., Inhibitors of HIV nucleocapsid protein zinc fingers as candidates for the treatment of AIDS. *Science* **1995,** 270, (5239), 1194-1197.
94. Tummino, P. J.; Scholten, J. D.; Harvey, P. J.; Holler, T. P.; Maloney, L.; Gogliotti, R.; Domagala, J.; Hupe, D., The in vitro ejection of zinc from human immunodeficiency virus (HIV) type 1 nucleocapsid protein by disulfide benzamides with cellular anti-HIV activity. *Proc. Natl. Acad. Sci. USA* **1996,** 93, (3), 969-973.
95. Rice, W. G.; Turpin, J. A.; Huang, M.; Clanton, D.; Buckheit, R. W., Jr.; Covell, D. G.; Wallqvist, A.; McDonnell, N. B.; DeGuzman, R. N.; Summers, M. F.; Zalkow, L.; Bader, J. P.; Haugwitz, R. D.; Sausville, E. A., Azodicarbonamide inhibits HIV-1 replication by targeting the nucleocapsid protein. *Nat. Med.* **1997,** 3, (3), 341-345.
96. Harada, Y.; Sakurada, K.; Aoki, T.; Thomas, D. D.; Yanagida, T., Mechanochemical coupling in actomyosin energy transduction studied by in vitro movement assay. *J. Mol. Biol.* **1990,** 216, (1), 49-68.

97. King, D. S.; Fields, C. G.; Fields, G. B., A cleavage method which minimizes side reactions following Fmoc solid phase peptide synthesis. *Int. J. Peptide Protein Res.* **1990,** 26, 255-266.
98. Laurence, T. A.; Kong, X.; ger, M. J.; Weiss, S., Probing structural heterogeneities and fluctuations of nucleic acids and denatured proteins. *Proc. Natl. Acad. Sci. USA* **2005,** 102, (48), 17348-17353.
99. Ying, L.; Wallace, M. I.; Balasubramanian, S.; Klenerman, D., Ratiometric Analysis of Single-Molecule Fluorescence Resonance Energy Transfer Using Logical Combinations of Threshold Criteria: A Study of 12-mer DNA. *J. Phys. Chem. B* **2000,** 104, 5171-5178.
100. Hong, M. K.; Harbron, E. J.; O'Connor, D. B.; Guo, J.; Barbara, P. F.; Levin, J. G.; Musier-Forsyth, K., Nucleic acid conformational changes essential for HIV-1 nucleocapsid protein-mediated inhibition of self-priming in minus-strand transfer. *J. Mol. Biol.* **2003,** 325, (1), 1-10.
101. Guo, J.; Wu, T.; Kane, B. F.; Johnson, D. J.; Henderson, L. E.; Gorelick, R. J.; Levin, J. G., Subtle alterations of the native zinc finger structures have dramatic effects on the nucleic acid chaperone activity of Human Immunodeficiency Virus Type 1 nucleocapsid protein. *J. Virol.* **2002,** 76, 4370-4378.
102. Cantor, C. R.; Schimmel, P. R., *Biophysical Chemistry. Part III. The Behavior of Biological Macromolecules*. W. H. Freeman & Co.: San Francisco, 1980; p 849-1371.
103. Misra, V. K.; Draper, D. E., On the role of magnesium ions in RNA stability. *Biopolymers* **1998,** 48, (2-3), 113-135.
104. Widengren, J.; Schwille, P., Characterization of photoinduced isomerization and back-isomerization of the cyanine dye Cy5 by fluorescence correlation spectroscopy. *J. Phys. Chem. A* **2000,** 104, (27), 6416-6428.
105. Widengren, J.; Schweinberger, E.; Berger, S.; Seidel, C. A. M., Two new concepts to measure fluorescence resonance energy transfer via fluorescence correlation spectroscopy: Theory and experimental realizations. *J. Phys. Chem. A* **2001,** 105, (28), 6851-6866.
106. Liu, H.-W.; Cosa, G.; Landes, C. F.; Zeng, Y.; Kovaleski, B. J.; Mullen, D. G.; Barany, G.; Musier-Forsyth, K.; Barbara, P. F., Single-Molecule FRET Studies of Important Intermediates in the Nucleocapsid-Protein-Chaperoned Minus-Strand Transfer Step in HIV-1 Reverse Transcription. *Biophys. J.* **2005,** 89, (5), 3470-3479.
107. Berkhout, B.; Vastenhouw, N. L.; Klasens, B. I. F.; Huthoff, H., Structural features in the HIV-1 repeat region facilitate strand transfer during reverse transcription. *RNA* **2001,** 7, (8), 1097-1114.
108. Takahashi, K. I.; Baba, S.; Chattopadhyay, P.; Koyanagi, Y.; Yamamoto, N.; Takaku, H.; Kawai, G., Structural requirement for the two-step dimerization of human immunodeficiency virus type 1 genome. *RNA* **2000,** 6, (1), 96-102.
109. Andersen, E. S.; Contera, S. A.; Knudsen, B.; Damgaard, C. K.; Besenbacher, F.; Kjems, J., Role of the Trans-activation Response Element in Dimerization of HIV-1 RNA. *J. Biol. Chem.* **2004,** 279, (21), 22243-22249.

110. Skripkin, E.; Paillart, J.; Marquet, R.; Ehresmann, B.; Ehresmann, C., Identification of the Primary Site of the Human Immunodeficiency Virus Type 1 RNA Dimerization in vitro. *Proc. Natl. Acad. Sci. USA* **1994,** 91, (11), 4945-4949.
111. Muriaux, D.; De Rocquigny, H.; Roques, B. P.; Paoletti, J., NCp7 activates HIV-1 Lai RNA dimerization by converting a transient loop-loop complex into a stable dimer. *J. Biol. Chem.* **1996,** 271, (52), 33686-33692.
112. Rist, M. J.; Marino, J. P., Mechanism of nucleocapsid protein catalyzed structural isomerization of the dimerization initiation site of HIV-1. *Biochemistry* **2002,** 41, (50), 14762-14770.
113. Chang, K. Y.; Tinoco, I., The structure of an RNA "kissing" hairpin complex of the HIV TAR hairpin loop and its complement. *J. Mole. Biol.* **1997,** 269, (1), 52-66.
114. Comolli, L. R.; Pelton, J. G.; Tinoco, I., Mapping of a protein-RNA kissing hairpin interface: Rom and Tar-Tar. *Nucleic Acids Res.* **1998,** 26, (20), 4688-4695.
115. Chang, K.; Tinoco, I., Jr, Characterization of a "Kissing" Hairpin Complex Derived from the Human Immunodeficiency Virus Genome. *Proc. Natl. Acad. Sci. USA* **1994,** 91, (18), 8705-8709.
116. Lapadat-Tapolsky, M.; Pernelle, C.; Borie, C.; Darlix, J.-L., Analysis of the nucleic acid annealing activities of nucleocapsid protein from HIV-1. *Nucleic Acids Res.* **1995,** 23, (13), 2434-2441.
117. Baba, Y.; Kagemoto, A., Influence of magnesium ions on helix-coil transition of DNA determined by modified differential scanning calorimeter *Biopolymers* **2004,** 13, (2), 339-334.
118. Zuker, M., Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **2003,** 31, (13), 3406-3415.
119. Zuker, M.; Mathews, D. H.; Turner, D. H., Algorithms and thermodynamics for RNA secondary structure prediction: A practical guide. *Kluwer Academic Publishers: Dordrecht*, 1999; p 11-43.
120. SantaLucia, J., Jr., A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA* **1998,** 95, (4), 1460-1465.
121. Peyret, N. Prediction of nucleic acid hybridization parameters and algorithms. *Wayne State University, Detroit*, 2000.
122. Haddrick, M.; Lear, A. L.; Cann, A. J.; Heaphy, S., Evidence that a Kissing Loop Structure Facilitates Genomic RNA Dimerisation in HIV-1. *J. Mole. Biol.* **1996,** 259, (1), 58-68.
123. Paillart, J.-C.; Skripkin, E.; Ehresmann, B.; Ehresmann, C.; Marquet, R., A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proc. Natl. Acad. Sci. USA* **1996,** 93, (11), 5572-5577.
124. Paillart, J.-C.; Westhof, E.; Ehresmann, C.; Ehresmann, B.; Marquet, R., Non-canonical interactions in a kissing loop complex: the dimerization initiation site of HIV-1 genomic RNA. *J. Mole. Biol.* **1997,** 270, (1), 36-49.
125. Godet, J.; de Rocquigny, H.; Raja, C.; Glasser, N.; Ficheux, D.; Darlix, J.-L.; Mely, Y., During the Early Phase of HIV-1 DNA Synthesis, Nucleocapsid Protein Directs Hybridization of the TAR Complementary Sequences via the Ends of their Double-stranded Stem. *J. Mole. Biol.* **2006,** 356, (5), 1180-1192.

126. Vuilleumier, C.; Bombarda, E.; Morellet, N.; Gérard, D.; Roques, B. P.; Mély, Y., Nucleic acid sequence discrimination by the HIV-1 nucleocapsid protein NCp7: a fluorescence study. *Biochemistry* **1999,** 38, (51), 16816-16825.
127. Paoletti, A. C.; Shubsda, M. F.; Hudson, B. S.; Borer, P. N., Affinities of the Nucleocapsid Protein for Variants of SL3 RNA in HIV-1. *Biochemistry* **2002,** 41, (51), 15423-15428
128. .Maki, A. H.; Ozarowski, A.; Misra, A.; Urbaneja, M. A.; Casas-Finet, J. R., Phosphorescence and optically detected magnetic resonance of HIV-1 nucleocapsid protein complexes with stem-loop sequences of the genomic Psi-recognition element. *Biochemistry* **2001,** 40, (5), 1403-1412.
129. McKinney, S. A.; Declais, A.-C.; Lilley, D. M. J.; Ha, T., Structural dynamics of individual Holliday junctions. *Nature Struct. Biol.* **2003,** 10, (2), 93-97.
130. Cosa, G.; Zeng, Y.; Liu, H.-W.; Landes, C. F.; Makarov, D. E.; Musier-Forsyth, K.; Barbara, P. F., Evidence for Non-Two-State Kinetics in the Nucleocapsid Protein Chaperoned Opening of DNA Hairpins. *J. Phys. Chem. B* **2006,** 110, (5), 2419-2426.
131. Russell, R.; Zhuang, X.; Babcock, H. P.; Millett, I. S.; Doniach, S.; Chu, S.; Herschlag, D., Exploring the folding landscape of a structured RNA. *Proc. Natl. Acad. Sci. USA* **2002,** 99, (1), 155-160.
132. Russell, R.; Herschlag, D., Probing the folding landscape of the Tetrahymena ribozyme: commitment to form the native conformation is late in the folding pathway. *J. Mole. Biol.* **2001,** 308, (5), 839-851.
133. Kanevsky, I.; Chaminade, F.; Ficheux, D.; Moumen, A.; Gorelick, R.; Negroni, M.; Darlix, J.-L.; Fosse, P., Specific Interactions Between HIV-1 Nucleocapsid Protein and the TAR Element. *J. Mole. Biol.* **2005,** 348, (5), 1059-1077.
134. Manning, G. S., On the application of polyelectrolyte limiting laws to the helix-coil transition of DNA. V. Ionic effects on renaturation kinetics. *Biopolymers* **1975,** 15, 1333-1343.
135. Frank-Kamenetskii, M. D.; Anshelevich, A. A.; Lukashin, A. V., Polyelectrolyte model of DNA. *Sov. Phys. Uspekhi* **1987,** 151, 595-618.
136. Rouzina, I.; Bloomfield, V. A., Heat capacity effects on the melting of DNA. 1. General aspects. *Biophys. J.* **1999,** 77, (6), 3242-3251.
137. Hargittai, M. R. S.; Gorelick, R. J.; Rouzina, I.; Musier-Forsyth, K., Mechanistic Insights into the Kinetics of HIV-1 Nucleocapsid Protein-facilitated tRNA Annealing to the Primer Binding Site. *J. Mole. Biol.* **2004,** 337, (4), 951-968.
138. Gabus, C.; Derrington, E.; Leblanc, P.; Chnaiderman, J.; Dormont, D.; Swietnicki, W.; Morillas, M.; Surewicz, W. K.; Marc, D.; Nandi, P.; Darlix, J.-L., The prion protein has RNA binding and chaperoning properties characteristic of nucleocapsid protein NCp7 of HIV-1. *J. Biol. Chem.* **2001,** 276, (22), 19301-19309.
139. Liu, H.-W.; Zeng, Y.; Landes, C. F.; Kim, Y. J.; Zhu, Y.; Ma, X.; Vo, M.-N.; Musier-Forsyth, K.; Barbara, P. F., Insights on the role of nucleic acid/protein interactions in chaperoned nucleic acid rearrangements of HIV-1 reverse transcription. *Proc. Natl. Acad. Sci. USA* **2007,** 104, (13), 5261-5267.
140. Portman, D. S.; Dreyfuss, G., RNA annealing activities in HeLa nuclei. *The EMBO J.* **1994,** 13, (1), 213-221.

141. Moumen, A.; Polomack, L.; Roques, B.; Buc, H.; Negroni, M., The HIV-1 repeated sequence R as a robust hot-spot for copy-choice recombination. *Nucleic Acids Res.* **2001,** 29, (18), 3814-3821.

142. Li, P. T. X.; Collin, D.; Smith, S. B.; Bustamante, C.; Tinoco, I., Jr., Probing the Mechanical Folding Kinetics of TAR RNA by Hopping, Force-Jump, and Force-Ramp Methods. *Biophys. J.* **2006,** 90, (1), 250-260.

Vita

Hsiao-Wei Liu was born in Taipei, Taiwan on June 7th, 1974, the second child of Chung-Tien Liu and Yin-Lien Hsun. After graduating from Lan-Yang Girls' High School, I-Lan, Taiwan, in 1992, she attended the National Chung-Cheng University majoring in Chemistry with a minor in Education. During her undergraduate education, she did her undergraduate research with Professor Bor-Cherng Hong working on the organometallic syntheses. In June 1997, she graduated with a Bachelor of Science degree in Chemistry. For the following year, she did her teaching internship in Chia-Yi Junior High School, Chia-Yi, Taiwan. Subsequently, she became a certified high school teacher in Chemistry and Physics and continued her teaching career in Huey-Deng High School, I-Lan, Taiwan. In the summer 2001, she joined the Professor Pi-Tai Chou's group in National Taiwan University, Taipei as a research assistant working on developing the molecular probe for the biological molecule detection. In August 2002, she enrolled in Graduate School at the University of Texas at Austin under the supervision of Professor Paul F. Barbara. During her Ph.D. studies, her work resulted in six peer reviewed publications in high impact journals including *PNAS*, *JACS*, *Biophysical J.* and *J Phys. Chem.*. In March 2005 and December 2006, she was awarded the Rom Rhome Award for Excellence in Material Science and Rohm-Hass Professional Development Award, respectively. In May 2007, she was awarded the Welch Fellowship Award.

Permanent address: No. 40 Sec 3 Fu-hsing Rd., I-Lan City 260, Taiwan

This dissertation was typed by the author.